

Document Version

Final published version

Licence

CC BY

Citation (APA)

van Donge, W. (2026). *Public Information Infrastructure Stewardship: A study of PII Quality, Design Variables, and Prevailing Configurations Dissertation*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:9f002ee5-e6d4-46ca-bea5-39e4601be3f5>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

The background is a solid teal color with a complex, white, geometric pattern of interconnected lines and shapes, resembling a stylized maze or circuit board. The pattern consists of various rectangular and irregular shapes with rounded corners, some of which are nested within others, creating a sense of depth and complexity.

*Public Information
Infrastructure
Stewardship*

WENDY
VAN DONGE

Public Information Infrastructure
Stewardship

Public Information Infrastructure Stewardship

A study of PII Quality, Design Variables, and Prevailing
Configurations

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus,
Prof.dr.ir. H. Bijl,
Chair of the Board for Doctorates
to be defended publicly on
Thursday 28 May 2026 at 15:00

By

Wendy VAN DONGE

This dissertation has been approved by the promotor.

Promotor: Prof. dr. ir. N. Bharosa

Promotor: Prof. dr. ir. M.F.W.H.A. Janssen

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. N. Bharosa,	Delft University of Technology, promotor
Prof. dr. ir. M.F.W.H.A. Janssen,	Delft University of Technology, promotor

Independent Members:

Prof. dr. ir. G.A. de Reuver,	Delft University of Technology
Prof. dr. T.A.P. Metze,	Delft University of Technology
Prof. mr. dr. M.H.A.F. Lokin,	Open University
Prof. dr. A.G. Plugge,	Nyenrode Business Universiteit
Dr. A.M.G. Zuiderwijk- van Eijk	Delft University of Technology



Logius
Ministerie van Binnenlandse Zaken en
Koninkrijksrelaties

DIGICAMPUS

Keywords: Public information infrastructures, Stewardship, Configurational analysis,
Data sharing, PII lifecycle, Design choices.

Printed by: Ridderprint.nl

Cover by: Ben Sminia & Sila Akçakoca

ISBN 978-94-6518-298-8

An electronic version of this dissertation is available at: <http://repository.tudelft.nl>

*"It's a dangerous business, Frodo, going out your door. You step onto the road, and if you don't keep your feet,
there's no knowing where you might be swept off to."*

- from The Lord of the Rings, written by J.R.R. Tolkien

Table of Contents

Table of Contents	vii
Acknowledgements	xi
1 Introduction	1
1.1 Introduction	2
1.2 Complexity and configurational dynamics	4
1.3 Finding configurations of PII	9
1.4 Research questions	12
1.5 Knowledge gaps and contributions	13
1.6 Thesis outline	16
2 Research approach	19
2.1 Mixed-method research approach	22
2.2 Validity and reliability	24
2.3 Research phase 1: <i>Definitions</i>	28
2.4 Research phase 2: <i>Design variables</i>	32
2.5 Research phase 3: <i>Configurations</i>	36
2.6 Research phase 4: <i>Stewardship</i>	44
2.7 Chapter conclusion	45
3 Theoretical background	47
3.1 Public information infrastructures	48
3.2 Defining PII stewardship	57
3.3 Quality framework for PII	63
3.4 Chapter conclusion	68
4 Cases of Public Information Infrastructures	71
4.1 Explorative cases	72
4.2 Descriptive cases	76
4.3 Cross-case contextual factors	78
4.4 Chapter conclusion	81
5 Design variables of a PII	83

5.1	Data design variables in a PII.....	85
5.2	Technical design variables in a PII.....	90
5.3	Governance design variables in a PII.....	98
5.4	Chapter conclusion.....	104
6	Configurations	107
6.1	Data sub-configurations.....	109
6.2	Technical sub-configurations.....	117
6.3	Governance sub-configurations.....	129
6.4	Zooming in on a single governance structure within a governance sub-configuration.....	140
6.5	Prevailing PII configurations.....	143
6.6	Quality of prevailing configurations.....	148
6.7	Chapter conclusion.....	156
7	Synthesis: Empirically derived classification of PII stewardship.....	159
7.1	Changes in PII configurations throughout its lifecycle.....	160
7.2	PII stewardship.....	163
7.3	Chapter conclusion.....	171
8	Conclusions and recommendations.....	175
8.1	Findings.....	176
8.2	Contributions.....	182
8.3	Reflection, Limitations and Further research.....	184
8.4	Limitations.....	187
8.5	Future Research.....	188
	References	191
	Appendices.....	203
	Appendix A: Table SLR.....	203
	Appendix B: Case descriptions.....	205
	Summary	213
	Samenvatting.....	219
	Curriculum Vitae.....	225

Acknowledgements

The selected header quote is a line from *The Lord of the Rings*, written by J. R. R. Tolkien. It speaks about life's unpredictability outside of our comfort zone. The past five years have been a rollercoaster of highs and lows. In my view, completing this journey and earning a PhD can only be achieved through perseverance and having the right people around to keep you grounded. Writing this dissertation signifies the end of a long, intense, and educational journey. It is a journey I have never travelled alone. This work, therefore, stands on the unconditional support, love, and trust of the people around me.

For those I love, and whose love has shaped me.

To my love, Ben, your patience and sense of humour have been the quiet force behind this work. In my chaos, you are my calm. Thank you for filling even the hardest days with laughter. To my loving parents, Edwin and Diny: If it were not for you, none of this would have happened, and I am truly grateful. Your unconditional love and support have guided me through every stage of this journey. From our heated discussions that kept me sharp, to the shared meals and the quiet comfort of doing absolutely nothing together. My dearest Danine, you are my lobster, my cheerleader, my drinking buddy and my best friend. When I need it the most, you are the ground beneath my feet.

There are friends who understand your mind, and friends who hold your heart. I am lucky to have both.

Special thanks to Amy: if anyone understands me, it is you. To Gitana, for the cosy family weekends. To Amber, thank you for the calls and support during our lunch breaks. To Dennis, Yvette, and Joey for the Monday evenings. To Eline, thank you for the endless stream of supportive messages. To Charlotte and Anne, thank you for dragging me to the Gym. To Wendy and Marco, for the endless number of Star Wars memes. To the ladies of Sofishticated, we may be different, but together we make one unforgettable group. Cheers to everyone who has reminded me to have a life outside of work.

Working together enables us to gain knowledge that we could never have acquired on our own.

I would also like to express my deepest gratitude to Logius, its employees, and my colleagues. I am extremely grateful for the support that enabled me to embark on this journey. Thanks to their support, I was able to combine work and study, and I would also like to thank Digicampus and Delft University of Technology for giving me the opportunity to start and complete my PhD project. To Louise, with whom I have had the pleasure of working during the countless interviews. To all the interview respondents, without whose participation, none of this would have been possible. Special thanks go to my (former) colleagues Edwin, Harry, Mattijn, Victor, Zoë, Wim, Han, Rob, Sylvia, Lizzy, Sander, and Maarten. Thank you for your support, involvement, substantive discussions, and feedback.

Some people inspire, some challenge, some accompany — and at times, one person does all three.

Finally, I would like to express my deepest gratitude to my supervisors. Professor Nitesh Bharosa, whose enthusiasm for the subject is truly infectious. Your deep love for the field shines through in everything you

do, and it inspires everyone around you! Professor Marijn Janssen, it is evident that your academic rigour and deep understanding of the field are invaluable to this manuscript. You have a remarkable talent for identifying weaknesses in the research and providing guidance on how to strengthen it. You do this not by criticising me, but by pushing for clarity and precision with the greatest kindness. Thank you both for guiding me on this journey.

Chapter 1

INTRODUCTION

1.1 INTRODUCTION

The ability of public organisations to generate societal value increasingly depends on how they collect and use data. Due to digitalisation, the volume of data has expanded rapidly. Over the past decades, as paper-based forms and processes have been converted into digital databases and automated workflows, government agencies have become large data organisations (Janssen & Kuk, 2016; Klievink et al., 2017). In the digital age, data has become a strategic asset for public agencies seeking to implement policies more effectively, detect societal trends earlier, and improve the efficiency of public service delivery. Inspired by the high levels of personalisation and user-friendliness achieved by leading Big Tech companies such as Google, Amazon, Facebook, Baidu, and Tencent, tech-savvy citizens, politicians, and policymakers now have higher expectations for public service delivery (United Nations, 2018). Government agencies face increasing pressure to operate more efficiently and to lessen the burden on taxpayers (Janssen & Estevez, 2013).

By leveraging the constant stream of new information-sharing technologies and integrating multiple information sources, governments can act proactively rather than reactively. In practice, we have seen that data sharing enables the automatic identification of parents at border control, ensuring that children can travel safely with the correct guardians. In healthcare, information sharing allows for a more comprehensive picture of a person's situation and enables better, more targeted care. Information sharing also plays a crucial role in the labour market, linking job vacancies to the CVs of individuals who are not currently in the labour market, helping them find work more quickly and receive better support.

Although the demand for data to support public processes such as policymaking, inspection, enforcement, and public service delivery is increasing, governments do not always have direct access to the data they require. Much of the relevant data is held by other parties, including citizens, private organisations, research institutions, civil society organisations, and international platforms. Moreover, information is often scattered across various public agencies, and information sharing can be impeded by the lack of legal grounds, purpose limitations and non-interoperable systems. This creates complexity for governmental organisations. While they have a public responsibility to perform public processes and provide high-quality services, they rely heavily on data and information flows originating outside their organisational boundaries. This necessitates the establishment of information infrastructures between these, often external, data sources and public organisations.

An information infrastructure is a combination of information and communication technologies that go beyond simple technology or system combinations (Hanseth & Monteiro, 1998). An information infrastructure describes the social and technical elements that enable information to be used for multiple purposes. In this research context, an information infrastructure refers to the elements that enable data to be created in one place and utilised elsewhere. It is the integration of digital systems and services that facilitates the flow of information as a resource across multiple organisations. According to Hanseth & Monteiro (1998), public organisations have played a large role in the development of information

infrastructures, for example, in health care. Furthermore, public organisations themselves require public information infrastructures to acquire the data necessary for their processes.

In literature, information infrastructures are referred to as the (digital) facilities and services necessary for performing human activities and can be compared to ‘enabling resources in network form’ (Bowker et al., 2010, p.2). Since public sector organisations often require information beyond their internal operations and business processes, they design public information infrastructures (PII) to obtain the necessary information. This research focuses specifically on public information infrastructures, as the term information infrastructures is too broad for addressing data exchange with public organisations. Additionally, this research examines twenty-two cases in which the information is required for the primary processes of the information receiver, as there is a significant dependence on the information in these instances. In these cases, a larger design space for government becomes apparent, which will be discussed in more detail together with the design dilemma in Chapter 1.2. The focus on governments as data recipients is merely a subset of PII, as public organisations can also be data suppliers to private organisations. A PII is a socio-technical system that is not static but constantly evolves based on the decisions of the stakeholders involved in the public. It consists of data, technical, and governance components, together forming a PII configuration. Chapter 3 provides more elaboration on what constitutes a PII. In some cases, public agencies are directed by law to ensure the availability of PII. They are consequently responsible for its design, implementation, and maintenance (Wet modernisering elektronisch bestuurlijk verkeer, 2023). However, research on the authorities involved in the configurations of a PII remains limited.

While in some cases information can be exchanged directly between sender and receiver, there is a growing trend towards the development of information infrastructures that involve multiple stakeholders in the data-sharing process (Bharosa et al., 2015; Klievink et al., 2016). Here, we observe that the new stakeholders involved are intermediary organisations that provide technical components such as systems, software, or services. In the public sector, however, additional stakeholders such as policymakers, standardisation bodies, and supervisory authorities may also be involved (Bharosa et al., 2015). Chapter 3 provides a more in-depth understanding of PII, and Chapter 4 presents 23 empirical cases of PII. These cases, ranging from business reporting and tax collection to e-invoicing and customs declarations, demonstrate how important PII have become for societal functioning. Three of these cases will be used to explore what design choices can be made in a PII, and what constitutes PII stewardship. Two of these cases are also used to descriptively determine the configurations found in practice. For this purpose, we added 20 cases to develop a well-established understanding of PII configurations and PII stewardship.

Although it has broad applicability and use, PII can be subject to inertia, meaning it tends to remain unchanged. Reasons for this vary and can include the need for stability and predictability (if it is not broken, do not fix it), low public and political interest or rewards, and a lack of investment in modernisation. However, the context and environment are dynamic, including a constant stream of new policies, laws, technologies, and actors in the domain. With the emergence of new laws (e.g., SDG, eIDAS, DSA, DMA), digital technologies (e.g., APIs, digital ledgers, wallets, AI), and actors in the field, renewed focus is required on the design and redesign of PII, as well as on the consequences of design choices for the actors and their

quality requirements (European parliament, 2018; European parliament, 2014; European parliament 2022a; European parliament 2022b). While extensive work has been done on information infrastructures, we know little about the different PII configurations found in practice. Furthermore, the literature offers limited guidance on how to steer PII and handle the numerous design choices that influence PII quality. Concepts such as governance are too broad to capture this role, while data stewardship is often too operational and confined to a specific organisation. Drawing on the concept of data stewardship, which refers to a set of tasks focused on ensuring that the right data reaches the correct processes in the appropriate format and complies with applicable laws and regulations (Dawes, 1996), this dissertation introduces the concept of PII stewardship. PII stewardship extends the principles of data stewardship to encompass an entire infrastructure, defining the collection of authorities for maintaining the quality of PII. The authorities within a PII are not always concentrated in a single place; instead, they can arise across different layers and governance structures. This study focuses on identifying specific authorities and parts of organisations that contribute to maintaining and safeguarding quality within the PII. PII stewardship must accommodate the distributed and interconnected nature of public information infrastructures. Authorities tasked with maintaining quality can therefore be shared among multiple organisations in different ways. This research conceptualises the role of PII stewardship, demonstrating that the tasks of PII stewards depend on the specific PII configuration. This stewardship must be exercised throughout the PII lifecycle, encompassing the design, implementation, expansion, and transition phases. In this research, we arrive at the following definition: PII stewardship refers to the authorities responsible for ensuring data, technical, and governance quality in designing, maintaining, and adapting a public information infrastructure throughout its lifecycle.

The definition includes several terms that will need additional clarification. The concept of PII stewardship will be further elaborated in Chapter 3.2. Information about the lifecycle of a PII will be provided in Chapter 3.1.5, while details on PII quality will be covered in Chapter 3.3.

1.2 COMPLEXITY AND CONFIGURATIONAL DYNAMICS

The design and governance of PII are often fragmented, context-dependent, and guided by implicit assumptions or ad hoc decisions. These infrastructures are no longer isolated or monolithic systems, but rather complex socio-technical arrangements. Organisations involved in designing and maintaining PII face the challenging task of aligning multiple layers of complexity: the technical and organisational challenges of designing an information infrastructure, the intricacies of engaging with private-sector actors, and the coordination required for effective collaboration with other public organisations. However, they encounter many issues in this alignment due to complexity and configurational dynamics.

II are inherently complex designs by their very nature. An information infrastructure is "*the basic information technologies and organisational structures, along with the related services and facilities necessary for an enterprise or industry to function.*" (Tilson et al., 2010, p. 748). Designing the technical layer of an information infrastructure is complex in itself. Various technologies can be used to connect systems, raising questions about standardisation, interoperability, flexibility, robustness, and the ability to support future advancements.

Information infrastructures are not merely technical systems; they are integrations of information systems, communication technology, and other (non-)technical components (Aanestad & Jensen, 2011, p. 162). This requires continuous alignment as information infrastructures are no longer isolated systems, but are “*shared, open, heterogeneous, and evolving socio-technical systems of Information Technology (IT) capabilities*” (Hanseth & Lyytinen, 2010). They are becoming more loosely coupled, dynamic, and distributed (Ryazanova et al., 2016), requiring coordination across institutional, technological, and sectoral boundaries. The social aspect of the PII also involves complexity, as multiple stakeholders with different requirements, visions for future development, and desired outcomes are involved. These differing goals often do not align, complicating decision-making processes within the PII. Aligning both the social and technical components of the PII presents a significant challenge.

1.2.1 DEVELOPMENTS ENHANCING COMPLEXITY

On top of the existing complexity of designing a PII, with all stakeholders connecting across multiple technologies and systems, three developments further complicate the design process. These include rapid technological evolution, increasing collaboration between public organisations, and private-sector involvement in PII.

The first development is that information infrastructures have rapidly evolved over the past decade. As new (technical) components become available, existing infrastructures must be adapted and reconfigured. This ongoing redesign of information infrastructures requires new resources, both technical and social, to integrate the new elements into the existing infrastructure (Bowker et al., 2010). Therefore, the configurations are constantly evolving and require alignment. Authorities involved in the maintenance, design, and redesign of a PII for information sharing must navigate this socio-technical complexity while also upholding public values such as transparency, auditability, accountability, and privacy within the infrastructure. Balancing these values is essential to maintaining public trust and ensuring that the infrastructure serves the broader societal interests.

A second development that adds complexity to public information infrastructures is the increasing collaboration among public organisations. As Tilson et al. (2010) explain, “*generativity adds unforeseen properties to information systems, such as the capability to recombine data sources and to semi-automatically generate, assemble, and redistribute content. The associated increase of metadata is also creating new industries, such as the search industry.*” (p. 750). This generative nature not only enables data reuse but also facilitates the integration of data exchange processes across various information requests. Public organisations often require similar data, and by aligning these needs, efficiency in data collection can be improved. This can substantially lessen the burden on data providers. This principle is commonly known as ‘report once, use many’. Furthermore, multiple services within the PII, such as authentication, data transformation, quality checks, and other data-related functions, tend to overlap across data-sharing processes. By collaborating and consolidating these services, organisations can reduce administrative workload and increase efficiency for both data providers and recipients. Research by Plugge, Nikou and Janssen (2022) shows that, for example, shared service centres

benefit from standardisation and modularisation, which strengthen interoperability. This also contributes to more successful collaboration between the public organisations using the shared service centres. However, collaboration can also create differing requirements and intended uses of systems and standards, making alignment challenging. This increases the complexity of data processes, as varying demands are placed on the data and multiple actors are involved as both senders and receivers.

The third development is private-sector involvement in PII (Bharosa, 2022). Beyond the complexity introduced by collaboration with the public sector, complexities arise from the increasing involvement of the private sector in public data exchange. Before exploiting information through data analytics, information must be collected, stored, and processed (Zhou et al., 2014). Here, both public and private organisations can provide services and components for the data exchange process. The technological change from analogue to digital has unlocked new opportunities for innovative service provision. *“Digitising has the potential to remove the tight couplings between information types and their storage, transmission, and processing technologies—potentially shattering the dominant service model and the stability of the industrial organisation.”* (Tilson et al., 2010, p. 749). Both public and private organisations have found ways to ease and optimise information exchange processes, providing services for those involved in the information infrastructure.

Given the enormous commercial potential resulting from high information exchange volumes in the public domain, there is an incentive for start-ups and enterprises to enter the ‘GovTech’ market. The rise of GovTech will have a major impact on public services (Filer, 2019). *“GovTech refers to socio-technical solutions – that are developed and operated by private organisations – intertwined with public sector components for facilitating processes in the public sector”* (Bharosa, 2022, p.1). Innovation is often hard to come from within government organisations, and GovTech can provide public service innovation from outside these organisations (Bharosa, 2022). It can be created, used, and reused in multiple domains. Examples include services that provide high assurance, such as eIDs for authentication and electronic document signing, as well as applications that assist citizens in applying for public housing and managing debt through financial planning. For the accurate and convenient use of such apps, private tech providers (often start-ups) need access to government services (e.g., citizen number authentication) and data sources (e.g., citizen registry, income registry). Not only can large existing organisations now be part of large information infrastructures, but GovTech also opens the field to small tech players, creating an alternative ecosystem (Bharosa, 2022).

Governmental organisations in PII can benefit from embracing GovTech solutions; however, there are also risks. Bharosa (2022), for instance, highlights the risks of increased public agency dependency on (foreign) GovTech providers in a global market that tends to reward consolidation and scale. Additionally, in some countries, such as the Netherlands, laws prohibit government agencies from competing with private enterprises (Dutch government, 2025). If there are solution providers in the market, the government must ensure a level playing field. In the Netherlands, government agencies, such as the Tax Office, Chamber of Commerce, Office of Statistics, Office of Education, and the Social Housing Office already rely on information system components (e.g., eID, data preparation software, data taxonomies, data reporting software, API, validation services.) that are provided, managed, and updated beyond their organisational boundaries.

1.2.2 COMPLEXITY IN DESIGNING PII

Making choices about collaboration, involving private parties, implementing new technological developments, and harmonising them all is a monumental undertaking. It is complex because numerous variables and environmental factors must be considered when designing a PII.

It is the responsibility of stakeholders involved in (parts of) the design of the PII to consider these developments and consequences. Here, a role emerges for PII stewards—authorities involved in the evolution of PII and maintaining its quality throughout its lifecycle. The tasks of a PII steward can vary between different PII configurations. Achieving alignment and designing a PII requires a high level of knowledge and experience. However, the PII stewards and stakeholders involved in PII design face the design dilemma, the paradox of change, and the paradox of control.

The design dilemma of information infrastructures relates to the tension between the desire to intentionally design and the impossibility of centrally controlling complex socio-technical systems (Hanseth & Lyytinen, 2010). As Star (1999) argues, information infrastructures are not static artefacts but are instead a constantly evolving network of technologies, practices, and actors. They are never entirely designed. PII are therefore also often not stable entities but evolve through ongoing interactions between social and technical components. The *evolutionary perspective* developed by Henfridsson and Bygstad (2013) offers valuable insights into how existing infrastructures can be adapted, scaled, and recombined over time. Answering the question of who is responsible for designing such infrastructures is complicated, because the design process is inherently dynamic, distributed and socio-technical, involving various stakeholders with different roles. Effective coordination in designing PIIs requires understanding the specific ‘design space’ of an infrastructure: which design variables can be decided upon, which configurations of these choices tend to prevail, and how these configurations relate to the quality of the PII. Although centralised steering is often unattainable, systematic insight into possible and prevailing configurations enables organisations and PII stewards to purposefully steer the infrastructure’s evolution and safeguard quality throughout its lifecycle.

The PII stewards and designers of the information infrastructure also have to consider the paradoxes of change and control. For the *paradox of change*, on the one hand, there is a need for a stable infrastructure to support the integration of new artefacts, processes, and actors. On the other hand, the infrastructure must remain flexible to accommodate ongoing growth and innovation. As Tilson et al. (2010) put it, “*digital infrastructures need to be stable to allow enrolment of new artefacts, processes, and actors; on the other hand, they must possess flexibility to allow unbounded growth.*” (p. 754). Despite the importance of this balance, there remains limited insight into the various design choices involved in integrating these elements into a PII.

In addition to managing change, government agencies must also navigate the paradox of control. Every organisation participating in an information infrastructure seeks a degree of autonomy or influence over it to shape system behaviour in its favour. These ‘points of control’ become critical levers in determining how the infrastructure functions. However, with many actors competing for influence and ongoing technological advancements, the infrastructure becomes a dynamic, contested space. Stakeholders continually attempt to

assert or protect points of control, often resulting in shifts in power and governance (Tilson et al., 2010). To some degree, autonomy or influence is important as public organisations depend on information to support (automated) processes and decision-making; the quality of that information is crucial. It directly affects the effectiveness of decisions and, consequently, the quality of public service delivery. At every stage, planning, implementing, executing, and enforcing policies, government organisations require high-quality data to perform their public duties. Poor data quality can lead to inefficiency and economic loss (Haug et al., 2011) as well as incorrect decisions that negatively affect citizens' lives (Jaeger & Bertot, 2010). Currently, there is no clear or shared understanding in academic literature of what constitutes 'quality' within public information infrastructures.

Currently, PII stewards and stakeholders involved in PII design lack a clear framework for their design space or insight into prevailing PII configurations in practice. Recognising this can help establish the PII and understand which configurations may lead to certain types of PII quality. This research, therefore, takes the first steps towards understanding and clarifying what constitutes PII quality and how PII stewardship and design choices in PII configurations can contribute to achieving it. Research is required to explore quality in these infrastructures, which are often fluid, interconnected, and inherently complex. Both in theory and practice, there is a lack of comprehension of these PII, making it challenging to predict the practical effects of particular design decisions, such as the distinction between centralised and decentralised infrastructure.

Information infrastructures are continually evolving and subject to ongoing change. Consequently, different phases of the infrastructure may require steering and decision-making. Gaining insight into these choices early can help public organisations make more informed, coherent decisions and ensure that infrastructure develops in line with public values and long-term objectives.

Furthermore, there is limited understanding of how these PII configurations can be maintained, designed and redesigned over time, especially in multi-actor environments with conflicting interests. Theoretical concepts such as governance and data stewardship alone are insufficient to fully grasp the authorities involved in maintaining, adapting, and evolving PII, which requires both practical data stewardship and high-level coordination aligned with governance. Here, we introduce a new concept of public information infrastructure stewardship, offering a promising perspective for exploring how PII can be designed, maintained, and adapted to enhance quality, long-term usability, and alignment with public values. There is a lack of conceptualisation and classification of PII stewardship in the current literature. PII stewardship refers to the authorities for ensuring data, technical, and governance quality in designing, maintaining, and adapting a public information infrastructure throughout its life cycles. However, there remains a gap in both the literature and practice regarding the connection between stewardship and the design configurations of PII. This research aims to address that gap.

Figure 1-4 offers an overview of the public information infrastructure and its related concepts. This figure shows that a PII configuration consists of data, technical, and governance design choices. The design choices will be explained in Chapter 5, and the (sub-)configurations of PII are explained in Chapter 6. For PII stewardship, we identified three empirically derived PII stewardship categories in this research: single-

actor stewardship, shared stewardship, and distributed stewardship. Chapter 7.2 provides further insights into these three empirically derived classifications.

The central premise of this research is that PII stewards are authorities for designing and managing PII configurations and ensuring their quality throughout the various stages of the design process. Since public information infrastructures are inherently distributed and interconnected, stewardship must reflect this structure. As a result, the responsibilities for maintaining quality can be shared among multiple organisations in various arrangements. Authorities within a PII are often distributed across different layers and governance levels, rather than being concentrated in a single location. This study aims to identify the specific authorities and organisational components responsible for steering the PII, therefore ensuring quality within the PII.

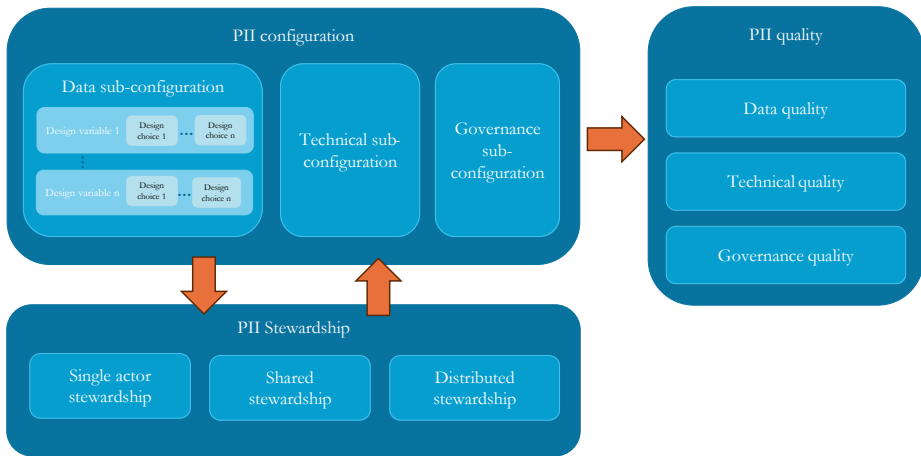


Figure 1-1. Simplified overview of PII concepts used in this study.

However, as the figure shows, the classification of PII stewardship is also shaped by the configuration of the PII. For example, if a PII configuration becomes more distributed, it is difficult for a single actor to serve as the sole authority for maintaining the entire PII. The PII configuration can be assessed against measures of PII quality, including data quality, technical quality, and governance quality. The quality framework is presented in Chapter 3.3, and the assessment of the configurations is presented in Chapter 6.6. This research focuses on identifying potential design choices within a PII and PII quality framework, exploring empirical configurations of PII design choices, understanding their relationship to PII quality, and research steering by PII stewards. First, we need to understand how PII manifests in design choices and configurations.

1.3 FINDING CONFIGURATIONS OF PII

As PII become increasingly central to societal functioning, understanding the underlying configuration of PII design choices is essential for effectively maintaining, adapting, and evolving high-quality PII. As

mentioned earlier, understanding the impact and outcomes of the design choices made by PII stewards and stakeholders can help them better grasp the consequences of these choices, their influence on the wider PII, and the design space available to them or potentially the limitations of that space. While previous work (e.g. Bharosa et al., 2015 and Klievink et al., 2016) provides some details on PII configurations, we lack a more systematic understanding of configurations of design variables in a PII. A configuration of a PII consists of choices made regarding the data exchanged within the PII, the technical components that facilitate the exchange, and the governance structures that coordinate, manage, and oversee (parts of) the infrastructure. The starting point for this research is that there must be several configurations of PII in practice, each with distinct characteristics shaped by design choices and context. We followed the configurational approach to examine the prevalent configurations of design variables of a PII identified in practice. The configurational approach allows for a deeper understanding of the complex relationships commonly found in practice. Instead of assuming that the effect of a single design choice on a PII outcome is isolated, this approach recognises that outcomes often result from multiple interacting design and contextual factors.

The configurational approach originates from contingency theory (Meyer et al, 1993). In this approach, research assumes that there is no single best way to organise or make decisions, as these depend on internal and external factors (Meyer et al., 1993). The approach explores unidirectional and linear relationships among attributes, in our case, the design variables. A design variable with its associated design choices represents a possible variation across different PIIs. Here, we observe that alternatives can be selected, resulting in different configurations.

A design variable represents a dimension (i.e. data ,technical, or governance) along which options can be selected, optimised, and implemented, as shown in Figure 1-1. A PII design variable must satisfy the following criteria:

First of all, a design variable involves an PII dimension (i.e. data, technical or governance) and options within that dimension. Based on the II literature, we conclude in Chapter 3.1 that a PII comprises data, technical design, and governance structures. We observe that the design of each of these dimensions consists of design variables, which together form a PII design.

Secondly, a design variable can be consciously chosen or configured by stakeholders. Next, a (configuration of) design variable(s) influences the infrastructure and its properties, such as safety, performance, and interoperability. Fourth, a design variable is measurable. They should be classifiable into options so that the differences between the designs are clear. For example, the organisation of the technical infrastructure could be: (1) central system, (2) semi-central, (3) decentralised, (4) distributed, or (5) hybrid. Finally, a design variable is relatively independent of other variables, not forcing another variable to change along with it. Combinations of design variables may occur more frequently, we label these combinations as configurations. These relationships between variables are not necessarily linear, as they may relate differently or not at all depending on the situation (Meyer et al., 1993). In this study, we anticipate that variables within a PII are not always linear. This is because design decisions in one area of the infrastructure can indirectly or gradually impact other areas, as the technical, data, and governance components interact and can either reinforce or constrain each other's outcomes. Some variables are likely to be related, while others can be

selected independently. Since a PII comprises many variables, we expect that some key choices will significantly influence its configuration.

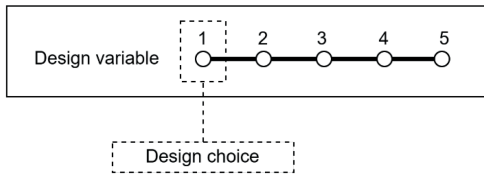


Figure 1-2. Relation design choice - design variable

In certain contexts, it becomes possible to identify configurations or clusters of variables. As Miller & Friesen (1977) noted: “It began to seem possible that there were a limited number of such models which could provide valuable insights into organisational/administrative behaviour in different environments.” (p. 255). When adopting the configurational perspective, the number of possible configurations is not unlimited. Although in theory, an infinite combination of variables is possible, in practice, only a few combinations are viable and will be present (Demers, 2007).

The patterns occur because we expect the variables in the configurations to be interdependent and to change intermittently or discretely (Meyer et al., 1993). The full configuration represents a complete set of design choices, as presented in Figure 1-3. Depending on the design choices, different configurations will arise. PII can have different or similar configurations, yet we anticipate that a few prevailing configurations will be found in practice, as presented in Figure 1-4.

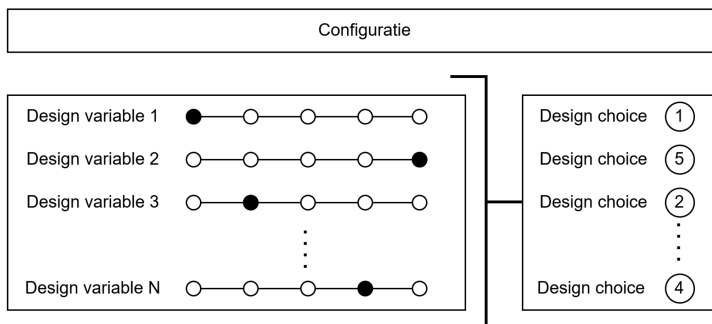


Figure 1-3. Configuration existing out of design choices.

A central focus of the analysis is identifying prevailing configurations as presented in Figure 1-4. A prevailing configuration is a pattern of design choices that either occurs frequently across cases or deviates from other configurations in an analytically relevant way. Such configurations often reflect underlying interdependencies between specific design choices, combinations that work particularly well together, or they may indicate best practices within a specific domain.

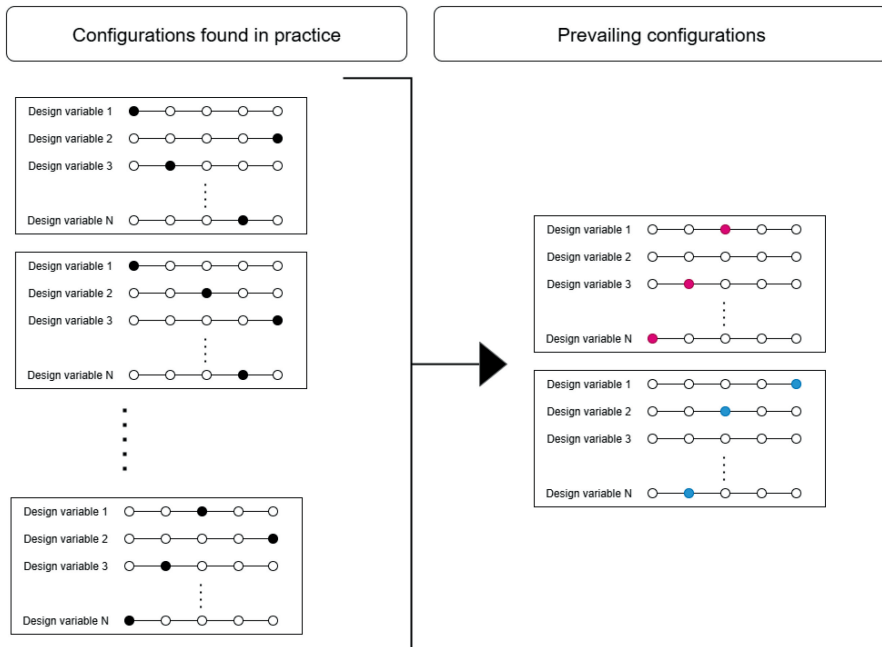


Figure 1-4. Relation between found configurations and prevailing configurations

Although the design space does not always encompass the full configuration of a PII, it remains essential to understand both the design variables and the prevailing configurations. This helps in understanding the range of choices available at the PII level and identifies which configurations manifest in practice. The configurations reveal how variables interact, facilitate the prospect of potential outcomes in terms of PII quality, and clarify the contextual factors influencing these configurations. In practice, this will help practitioners better understand the impact of their decisions and how these decisions affect other components within the PII, allowing them to anticipate potential trade-offs, dependencies, and implications for the stewardship and quality of the PII when making design decisions. This provides stakeholders in the PII with a clearer framework within their design space, and enables PII stewards and the authorities responsible for coordinating quality throughout the lifecycle to steer the PII.

1.4 RESEARCH QUESTIONS

In the context of a PII, multiple design choices can be made, leading to a specific configuration. This study aims to identify prevailing configurations of design choices and explore how they relate to different classifications of PII stewardship, based on the assumption that some configurations may be better suited for certain stewardship classifications. The main research question is two-fold: **What is PII stewardship, and how does PII stewardship relate to configurations of PII for data exchange with (semi-) public organisations?**

In the first phase of this research, titled *Definitions*, we focus on understanding PII and PII stewardship. We examine these concepts through a literature review. We also develop a framework to explore what PII quality entails, drawing on both the literature and an interview study. In this first phase, *Definitions*, we answer the following sub-research questions:

1. **What is a public information infrastructure?**
2. **How can public information infrastructure stewardship be defined?**

The second phase, titled *Design variables*, involves gaining insight into the various design variables that comprise PII configurations. We start with a literature review of potential design choices and design variables. Through a multiple-case study of three cases, we conduct interviews and document analysis to validate and develop a comprehensive framework of variables that represent PII design choices. Here, we answer the following sub-research question:

3. **What are the relevant design variables of a public information infrastructure?**

In the third phase, titled *Configurations*, this research employs a multiple-case study to conduct a cross-case analysis, examining how 22 cases are configured and identifying prevailing configurations of design choices. To identify PII configurations, we adopt a configurational approach. Additionally, we evaluate the strengths and weaknesses of PII configurations by answering the following sub-research questions:

4. **What are the prevailing configurations of public information infrastructures through which organisations exchange data with (semi-)public organisations?**
5. **What are the strengths and weaknesses of the prevailing configurations of public information infrastructures in terms of their quality?**

In the fourth and final research phase, titled *Stewardship*, we synthesise the results of the previous phases. Finally, a synthesis is conducted to bring the results together and generate insights into PII stewardship. This synthesis examines the identified PII configurations of the 22 cases in relation to the literature-based definition and characteristics of PII stewardship. By comparing configurations and the authorities acting as stewards, the analysis further refines the concept of PII stewardship and enables a classification of PII stewardship based on the empirical findings. In the final phase, we seek to answer the following sub-research question:

6. **What classification of public information infrastructure stewardship can be distilled from the findings of this study?**

Furthermore, this synthesis offers insights into the contextual factors by comparing the cases used in phases 2 and 3, helping us understand the conditions under which particular configurations arise and how these contexts influence the form and practice of stewardship. These findings are discussed in Chapter 4.3. By answering these six sub-questions, we answer the main research question in Chapter 8.

1.5 KNOWLEDGE GAPS AND CONTRIBUTIONS

This study started by addressing existing knowledge gaps in the literature and identifying the information deficiencies faced by practitioners in managing the increasing complexities of PII. Therefore, this study

contributes to both academic and practical discussions on the design and governance of PII. In the following sections, we will discuss the scientific knowledge gaps and contributions, as well as the societal relevance.

1.5.1 SCIENTIFIC KNOWLEDGE GAPS AND CONTRIBUTIONS

This research identifies four research areas with knowledge gaps. First, this thesis identified a knowledge gap in the research on public information infrastructure (PII). The literature often focuses on government and information sharing (e.g., Gil-García & Pardo, 2005; Layne & Lee, 2001; Linders, 2012), open government data (e.g., Dawes et al., 2016; Janssen et al., 2012; Purwanto et al., 2020), or the use of (big) data analytics within government (e.g., Janssen et al., 2017; Kim et al., 2014; Klievink et al., 2017). However, existing studies mainly focus on structuring information processes within a single organisation or between two organisations (Gil-García & Sayogo, 2016), often examining linear, one-directional exchanges. In contrast, PII involves more complex, multi-party information exchanges, where all actors collectively shape the infrastructure. This complexity highlights the need for deeper insights into the design, development, and governance of PII. By systematically identifying design variables and analysing prevailing configurations, this study bridges theory and practice. It offers a conceptual and empirical addition to the literature on information infrastructures, data governance, and digital public services, enabling scholars and practitioners to compare configurations, identify patterns, and derive insights for future design strategies.

Secondly, this research identified a gap in the literature on quality dimensions and their measures in PII. Although quality is frequently discussed in the information-sharing literature (e.g., Gil-García & Pardo, 2005), the specific dimensions and measures within a PII have not yet been clearly identified. Through empirical research, we explore which quality dimensions and measures are essential for shaping a PII and how certain design choices affect these measures. By pinpointing concrete quality dimensions and measures and linking them to configurations, this research not only clarifies theoretical concepts but also offers insights for future research on the design and management of PII.

Third, this is the first work to investigate PII stewardship. We state that PII stewardship refers to the authorities for ensuring data, technical, and governance quality in designing, maintaining, and adapting a public information infrastructure throughout its life cycles. These development stages are conceptualised based on the research (Bolton & Foxon, 2015; Edwards et al., 2009) and include the phases of ‘emergence and establishment’, ‘expansion and momentum’, ‘stagnation and inertia’, ‘transition and renewal’, which in turn evolve into ‘emergence and establishment’. In each of these phases, the PII steward strives for (different types of) PII quality.

The literature contains extensive research on data stewardship (e.g., Cheong & Chang, 2007; Dawes, 1996; Otto, 2011; Rosenbaum, 2010; Wendelborn et al., 2023). However, there remains a limited understanding of PII stewardship. Building on research related to Data stewardship (e.g., Dawes, 2010; Rosenbaum, 2010; Wende, 2007) and Governance (e.g., Khatri & Brown, 2010; Klievink & Janssen, 2014; Weill & Ross, 2005; Wende, 2007), this study is the first to establish the concept of PII stewardship. Specifically, it advances the concept of information infrastructure stewardship by connecting it to specific

required knowledge, life cycles, and design patterns observed in real-world cases. This marks the initial step towards understanding PII stewardship within the public sector.

Finally, there is no research yet towards identifying (sub-)configurations by using a configurational approach within the domain of public information infrastructures. The configuration method has not yet been utilised on this scale to identify configurations in the PII context. While hierarchical clustering is commonly used in fields such as medicine, biochemistry, engineering, and computer science, particularly in machine learning, its application in the public sector remains limited. Little to no research has employed this technique to cluster configurations of design choices in information infrastructures. This study highlights the broader potential of clustering methods in new domains, including public administration and information systems design.

1.5.2 SOCIETAL RELEVANCE

Governments worldwide are facing similar challenges related to data sharing, interoperability, and development (OECD, 2025). As public agencies become more reliant on data and socio-technical components beyond their direct or sole control, the ability of (semi-)public organisations to access, manage, and utilise data from external sources is vital for providing societal value. However, the configurations through which data exchange is organised are often shaped by implicit technical and institutional choices, without a clear understanding of their long-term effects. Shaping PII often occurs in complex contexts, and practitioners find it challenging to make the right choices and select configurations that result in high-quality PII. Decisions are frequently made without knowing their impact or their connection to other PII design variables. However, these choices have a substantial impact on the entire environment and are difficult to reverse. For practitioners such as policymakers, public managers, and infrastructure designers, this study contributes in two ways. First, this study offers a typology of design configurations and stewardship authorities that shape PII. Thinking in terms of PII configurations is new and marks a shift from thinking in terms of monolithic digital systems. Understanding the consequences of specific design choices enables public organisations to make more deliberate, context-aware decisions about how data infrastructures are constructed and maintained over time. It helps practitioners manage path dependencies, which restrict future options. Knowing the design choices in advance will help practitioners understand how to shape PII.

Second, this research offers insight into the empirically derived classification of PII stewardship, aiding in understanding the different roles of authorities involved in various aspects of PII. It helps practitioners to identify the resources and knowledge needed to manage PII effectively and recognise the boundaries and complexities of their duties. Ultimately, this research aims to promote more transparent, resilient, and value-aligned approaches to the development of PII in an increasingly complex and global information-sharing environment.

1.6 THESIS OUTLINE

This thesis is organised as follows. Chapter 2 offers a detailed description of the research methodology and the methods used. Chapter 3 presents the theoretical framework, reviewing the literature on PII and stewardship, and integrates empirical research on PII stewardship. Chapter 4 provides an overview of the empirical cases utilised in the research. It includes three exploratory cases aimed at gaining a deeper understanding of stewardship and PII; these cases are explained in more detail in this chapter. To identify the prevailing configuration, we analysed twenty-two descriptive cases and provided an overview in the second part of Chapter 4. Chapter 5 explores the design variables through a systematic literature review and multiple case studies. Using this list of design choices, the configuration of a PII can be defined.

In Chapter 6, we examine possible configurations identified in twenty-two cases. Here, we present data, governance, and technical sub-configurations. We identify common patterns in design choices across these cases, resulting in six prevailing configurations. Then, in Chapter 7, we distil the classifications of PII stewardship from this research and show how they relate to the configurations for data exchange with (semi-)public organisations. Finally, Chapter 8 concludes the thesis, showing its contribution, and outlining limitations and directions for future research.

Chapter 2

RESEARCH APPROACH

In the previous chapter, we established the need for research into PII and PII stewardship. This chapter expands on this by outlining the research approach used to deepen understanding in this field. It describes the chosen research design and the methods for data collection and analysis. This research aims to answer the main research question: **What is PII stewardship, and how does PII stewardship relate to configurations for data exchange with (semi-)public organisations?** To address this main question, six supporting research questions will be answered. Figure 2-1 provides an overview of the different research phases and their corresponding research questions. For each phase, it illustrates the required inputs, the research methods applied, and the outputs produced by each methodological step. In addition, the figure indicates which documentation is available in the repository to support these methods. References to the chapters in which these components are explained in more detail are also included.

This research is grounded in the critical realist approach (Bhaskar, 1975; Easton, 2010; Tsang et al., 1999), which argues that reality exists independently of our perceptions. Although this reality is independent of our perceptions, it cannot be fully accessed through research. Critical realism views reality as a layered, complex, and partly hidden structure. The empirical level can be observed and measured, whereas the actual and real levels are more concealed (Bhaskar, 1975). Because this reality is complex, partly open to interpretation, layered, and enacted through visible events that are expressions of underlying mechanisms and contextual conditions. Therefore, this study employs multiple methods to explore this reality. To understand it thoroughly, examining it across various cases is essential. This study examines 23 cases to gain a comprehensive understanding of PII in practice. Gaining a true understanding of the case requires a mixed methods design. The aim of this research, aligned with the critical realist approach, is not merely to describe observable patterns but also to identify the contextual factors that influence them. This study discusses these contextual factors in Chapter 4 and provides deeper insights into the configurations of design choices in Chapters 5 and 6.

As a result, this research employs the configurational approach to explain how specific combinations of design choices relate to PII stewardship roles and PII quality. Although the research aims to gain insights into the mechanisms, the explanations remain contextual. Our access to this reality is determined by the context.

Before explaining the four research phases, we will outline the research design and methods.

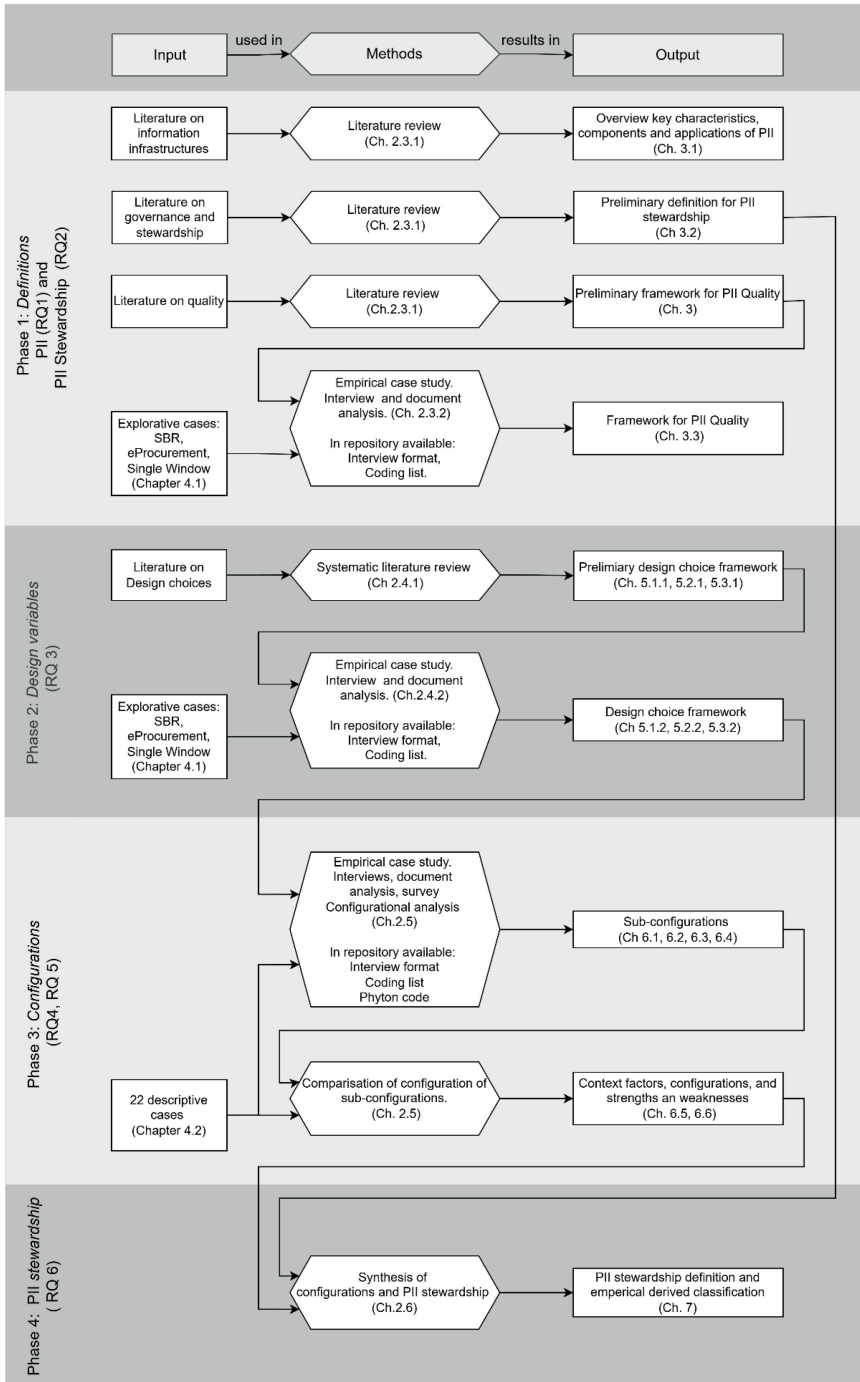


Figure 2-1. Overview of the research

2.1 MIXED-METHOD RESEARCH APPROACH

A mixed-methods research approach ensures that data collection is both detailed and thorough (Burke et al., 2004). A mixed-methods research approach allowed us to employ both quantitative and qualitative techniques. The aim of this approach is to gain a deeper, more comprehensive understanding of the research problem. The issue of configuring PII and the stewardship roles involved in this context is complex and multilayered, making a mixed-methods approach suitable, e.g., “*Mixed methods research offers powerful tools for investigating complex processes and systems*” (Fetters et al., 2013, p. 2134).

Using a mixed-methods research design aligns with the critical realism approach. As the approach states, the reality is complex and multi-layered, requiring multiple methods to understand this layered reality. The complex nature of reality necessitates diverse perspectives from both qualitative and quantitative research. Literature, existing case documentation, interviews, and surveys with respondents were used to gain insights into reality from multiple angles. “*Many research questions and combinations of questions are best and most fully answered through mixed research solutions.*” (Burke & Onwuegbuzie, 2004, p.18). Employing mixed methods, including both qualitative and quantitative approaches, facilitates triangulation, seeking confirmation of results across different data sources. Mixed methods aim to enhance generalisability and produce more substantial evidence to support research conclusions (Burke & Onwuegbuzie, 2004). They improve findings from one method with results from the other and help to offset the limitations inherent in individual methods (Brymann, 2006). “*A researcher can use the strengths of an additional method to overcome the weaknesses in another method by using both in a research study.*” (Burke et al., 2004, p.21).

Although there are some doubts about the ‘whatever works’ position that it is based on (e.g., Pawson & Tilley, 1997), combining both qualitative and quantitative research has become increasingly popular in recent years (Bryman, 2006). One disadvantage of using a mixed-method design is that it takes more time, as multiple methods must be learned and implemented, compared to a study with just one method (Burke & Onwuegbuzie, 2004). However, this does not outweigh the advantages of a mixed-method.

In the first two phases, as shown in Figure 2-1, we conducted a literature review and a systematic literature review (SLR) to build on previous research. In the literature review, we examined the field's literature to understand the concepts and identify key knowledge. Through the SLR, we establish the theoretical framework of design variables. In addition to the literature research, we utilise multiple case studies involving interviews, surveys, and document analysis. In the following subsections, we elaborate on the multiple case studies employed as part of the mixed-method approach in this research.

2.1.1 MULTIPLE CASE STUDIES

This chapter elaborates on the concept of a multiple case study and explains why this approach is suitable for the present research, as it will be applied in the subsequent chapters. Case studies are often used to investigate complex phenomena within their specific context, drawing on various sources to collect data (Yin, 2017). While the case study design follows Yin (2018), the analysis is guided by a critical realist perspective focusing on identifying underlying mechanisms. Case research is well-suited for critical realism.

It allows cases to be examined from multiple perspectives, fostering a broader understanding of the phenomenon. Case studies are especially useful for answering ‘how’ and ‘why’ questions. They mainly focus on researching the phenomenon in depth and can offer explanations of why or how it occurs. Another reason for selecting case studies is when it is not possible to manipulate the behaviour of those involved in the research. Unlike experiments, where researchers control events, case studies lack such control. This is why they often include an examination of contextual conditions in the study (Yin, 2017).

The cases for our research are selected on their similarity in being a PII, ensuring data sharing with governmental organisations, within the Dutch context. This research will analyse the individual design choices of PII and the quality of these PII, and compare cases to identify any patterns. Since multiple data sources are employed to shed light on a case, multiple case studies are often regarded as robust and reliable (Baxter & Jack, 2008).

Robust and reliable research is characterised by its consistency and reproducibility, meaning that applying the research repeatedly to the same object produces the same results each time (Babbie, 2020). By examining multiple cases, one draws on evidence from diverse contexts. This increases the likelihood that findings are not unique to a single instance. Systematically comparing and triangulating across cases allows us to confirm the observed patterns and configurations, thereby strengthening the validity of our conclusions. This approach ensures methodological rigour and enhances the generalisability and credibility of the research outcomes. In Chapter 2.2, we outline the techniques used to improve the research's validity and reliability.

2.1.2 DOCUMENT ANALYSIS

This chapter explains the concept of document analysis, which is used as a method in the later sections starting from Section 2.3. As presented in Figure 2-1, we utilised document analysis to answer RQ 2, 3, and 4. For document analysis, research is conducted using existing documentation. Primary documentation is defined as data that is not produced specifically for the research study (Sandra van Thiel, 2013). As noted by Riffe, Lacy & Fico (2019), documents are artefacts created by humans to record, communicate, and preserve different forms of knowledge, behaviours, and processes. The documents used in this study include a wide range of materials such as reports, agreements, and policies. It is clear that they also serve a communicative function, aimed at conveying messages. It is acknowledged that the text was not initially created as a research document; therefore, it may not fully align with established requirements. As a result, document analysis is often employed alongside other research methods to provide additional content and context. This research employs content analysis to examine the documents, which are analysed based on the different research topics relevant to each sub-research question. Document analysis is utilised in research phase two: *Design variables*, and phase three: *Configurations*.

2.1.3 SURVEY

This chapter explains the concept of surveys, which is used as a method in later sections starting from Section 2.5. We use surveys to support answering RQ 4 and 5. Using a survey provides a structured method for data collection, with predefined questions for respondents to answer. Surveys can gather both qualitative and quantitative data and enable researchers to systematically collect comparable data across samples (Babbie, 2010). A survey often aims to create codable responses (Presser et al., 2004). In our survey, we primarily use standardised response options to identify the design choices in each respondent's PII configuration. Previous research allowed us to identify relevant design choices, and the current survey aims to determine the specific configuration of these choices for each case. This approach allows for a systematic comparison of configurations across different PII cases. We included open-ended questions for elaboration on answers and provided space for respondents to reflect. To enhance the quality of the survey, we used pre-validated frameworks as input for the questions, pre-tested the survey, validated the results against existing documentation on the case, and conducted semi-structured interviews to focus on the responses provided in the survey.

2.1.4 SEMI-STRUCTURED INTERVIEWS

This chapter explains the concept of a semi-structured interview, which is used as a method in later sections starting from Section 2.3. As presented in Figure 2-1, this research utilises semi-structured interviews to answer RQs 2, 3, and 4. Interviews are a widely used method for qualitative research, especially within case studies (van Thiel, 2013). During an interview, the interviewer interacts directly with the respondent. Because of the presence of an interviewer, responses such as 'I do not know' or 'no answer' are often minimised, and the interviewer can ask follow-up questions to explore further (Babbie, 2010).

In this research, we use a semi-structured interview method. This method combines the structure of a predefined questionnaire with the flexibility to explore other topics in depth based on the participant's response. Semi-structured interviews are well-suited for exploratory studies because they allow for dialogue with respondents (Whiting, 2008). This method enables respondents to reflect on their own experiences, allowing new insights to emerge (Dearnley, 2005; Krauss et al., 2009). The semi-structured interviews are used to collect both subjective insights and as a triangulation for the surveys and document analysis. To strengthen internal validity and verify data accuracy, we also applied respondent validation, inviting interview respondents to review the interview draft and the case study report draft, and to confirm the accuracy of the identified data sources. The semi-structured interviews are conducted in research phase one: *Definitions*, research phase two: *Design variables*, and phase three: *Configurations*.

2.2 VALIDITY AND RELIABILITY

In research, ensuring the validity of findings is essential for establishing the credibility and applicability of the results. There are four tests for evaluating research quality: internal validity, construct validity, external validity, and reliability (Yin, 2017). Each plays a distinct role in enhancing the rigour of a study.

Internal validity determines if research can legitimately establish causal relationships, which is particularly relevant in explanatory research (Rowley, 2002). Triangulation of data, methods, and investigators helps reduce bias and strengthen the credibility of causal or explanatory findings. Additionally, pattern matching improves internal validity by directly comparing empirical results with theoretical expectations. Moreover, cross-case synthesis enhances internal validity when consistent patterns are observed across multiple cases.

In contrast, construct validity focuses on the extent to which a construct accurately reflects the intended concept. Achieving construct validity involves clearly defining the concept, aligning it with the study's objectives, and choosing suitable operational measures (Yin, 2017). Data triangulation improves construct validity by examining the same concept from multiple data sources. Additionally, methodological and investigator triangulation help ensure accurate capture of constructs by using different methods and perspectives. The pattern-matching technique verifies that constructs are consistent with their theoretical definitions and are interpreted uniformly.

The third type of validity is external validity, which relates to how well the study's findings can be generalised to other contexts. For studies involving multiple cases, applying replication logic, a process of testing whether similar results occur across different cases, enhances this type of validity (Rowley, 2002). Data triangulation enhances external validity by demonstrating that findings remain consistent across various contexts or sources. Additionally, cross-case synthesis emphasises which patterns can be generalised beyond individual cases.

Lastly, reliability refers to the consistency and reproducibility of results when applying the same methods under similar conditions. Reliability is enhanced when data, methodological, and investigator triangulation produce stable and consistent outcomes across sources, methods, and researchers. Pattern matching supports this by using a systematic and reproducible analysis process. Transparent documentation and cross-case synthesis further strengthen reliability by providing clear, reproducible procedures and an audit trail.

In the following subsections, we provide more details on the different techniques used to enhance the validity and reliability.

2.2.1 DATA TRIANGULATION

Data triangulation involves using multiple sources of evidence within a research study. This approach is especially recommended in case study research to strengthen the overall quality. In particular, data triangulation is often employed to improve construct validity (Yin, 2017). In this research, multiple data sources, including interviews, documents, and surveys, are used to address the research objectives thoroughly. Each of these sources contributed to the findings. Moreover, data triangulation enhances internal validity by ensuring that results are not reliant on a single source and are validated across multiple perspectives. For example, we examined PII from different angles, including quality, design perspectives, and stewardship roles. This offers a multifaceted and deeper understanding of a PII.

2.2.2 METHODOLOGICAL TRIANGULATION

Methodological triangulation involves using multiple methods, often both quantitative and qualitative, to address the same research question (Morse, 1991). Employing multiple methods can be done either simultaneously or sequentially. In the first approach, there is minimal interaction between the data sets, but the findings complement each other, whereas in the latter, the results of one method are crucial for planning the other (Morse, 1991). In this research, we utilise both.

In this study on PII and stewardship, methodological triangulation was employed through the combined use of literature reviews, interviews, document analysis, and surveys. Each method provided distinct insights. Interviews offered in-depth perspectives from stakeholders, document analysis captured formal and historical records, and surveys supplied broader quantifiable data. By integrating these methods, the research not only validated findings across different data collection strategies but also uncovered complementary dimensions of PII and stewardship.

This multi-method approach directly supported construct validity by aligning the operationalisation of concepts with multiple independent lines of evidence. It also strengthened internal validity by enabling cross-verification of findings, thereby ensuring that observed patterns were consistent across methods. Furthermore, methodological triangulation contributed to reliability by demonstrating that different data collection and analysis techniques produced converging results, which supports the reproducibility of the study.

2.2.3 INVESTIGATOR TRIANGULATION

In this study, we employ investigator triangulation to improve both validity and reliability. Bias may occur if interviews are not conducted impartially due to interactions between respondents and interviewers. It can also occur during the coding process, when researchers' interpretations may affect how the data are coded (Yin, 2017).

In the first three phases of the research, other researchers were invited to participate in researcher triangulation during the coding stages to ensure the reliability and validity of the study. The primary researcher first coded the interviews and documents. Afterwards, selected sections of the interviews and documents, along with the codebook, were shared with a second and third researcher. These researchers independently coded the assigned material using the codebook or proposed new codes if the existing ones were insufficient. In the third phase of the study, qualitative coding of the interviews and documents was carried out collaboratively by multiple researchers to increase the trustworthiness and reliability of this analytical stage. When discrepancies in coding arose between the first researcher and the second coder, they discussed them. Discrepancies mainly involved whether a code was a contextual factor or a deliberate choice. For the items discussed, consensus was reached and validated by an external researcher. Furthermore, the external researcher was peer debriefed throughout the process to further ensure rigour. The involvement of multiple researchers enhances reliability by improving consistency and objectivity through the recognition of similar results and by reducing dependence on a single researcher (Yin, 2017).

2.2.4 PATTERN MATCHING

To improve the rigour of this investigation, a pattern-matching methodology was employed. According to Yin (2017), pattern matching is a highly useful technique when conducting a case study analysis, as it compares empirically observed patterns with those predicted theoretically. Pattern matching was carried out during the second phase of the research, aligning literature-based design choices with empirical data. Additionally, the quality measures, grounded in existing literature, were validated by respondents within the cases.

This confirmed relevant design choices such as collaboration, the public-private distinction, and the centralisation debate. These have already been identified in the literature, highlighting the need for a more comprehensive understanding. The discrepancies uncovered allowed us to explore contextual nuances and alternative interpretations, deepening our insight into PII. By examining multiple cases, we identified recurring structures, moving beyond mere description to generate more profound insights into the underlying dynamics of PII. When interpreting findings on the configurations of PII, it is crucial to consider rival explanations. During the exploration of prevailing configurations, we examined all design choices to determine whether they contributed to the prevailing configuration. In the course of the research, we observed that some design choices, such as the push or pull type of infrastructure, which were deemed less relevant by respondents in research phase 2, became important in research phase 3. We did not identify additional design choices during the third phase of the research. For each sub-configuration, we clarify which variables correspond to the final configurations. The incorporation of multiple data sources and methodological approaches enabled us to cross-validate findings, thereby strengthening the foundation of this research. By explicitly engaging with these alternative explanations, we enhance the internal validity of our study and provide a more nuanced understanding of the complex dynamics underpinning PII.

2.2.5 CROSS-CASE SYNTHESIS

Aside from matching empirical patterns to theoretically predicted ones, we conducted a cross-case synthesis to identify within-case patterns and relationships across the cases (Yin, 2017). A cross-case synthesis can therefore only be performed in a multiple-case study research. In the cross-case synthesis, discussing differences and similarities between the cases is important. These differences should be reconsidered if they do not weaken the findings across the cases.

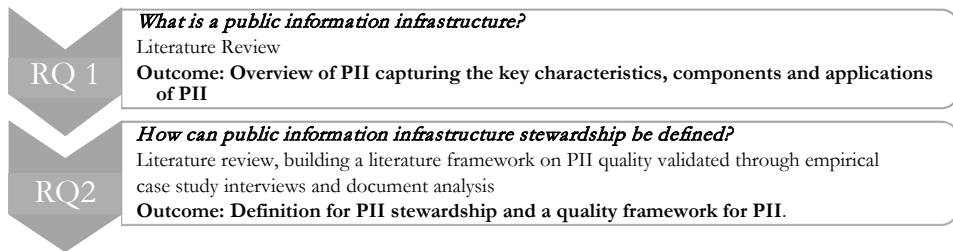
This study on PII used cross-case synthesis to explore recurring design choices and quality measures across multiple cases. Each case was initially analysed separately to understand its context and features. Then, the findings were compared to identify common patterns among the configurations. These differences and similarities were discussed with multiple researchers to reduce investor bias in the research. This process improved internal validity by showing that specific configurations and dynamics are not isolated events but occur regularly across various contexts. It also enhanced external validity by supporting analytical generalisation, illustrating how the findings might be applied to other contexts.

2.2.6 TRANSPARENT DOCUMENTATION

Reliability involves the reproduction of a study, leading to the same results. This can be achieved through thorough documentation of the research phases (Rowley, 2002). This thesis will be published in the TU Delft repository. Additionally, the interview questions, Python codes for hierarchical clustering, and the surveys used are available in the TU Delft repository on the DOI: [10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45](https://doi.org/10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45). In the following sections, we will explore the different research phases and how each method is used in each phase.

2.3 RESEARCH PHASE 1: *DEFINITIONS*

In the first phase of the research, the results of which are presented in Chapter 3, we aim to answer the first two sub-research questions. As shown in Figure 2-1 at the beginning of this chapter, we start with a literature review and conclude with empirical research towards quality in PII.



2.3.1 LITERATURE REVIEW

In this section of the research, we conduct a conceptual analysis, identifying and examining relevant theoretical sources to develop a conceptualisation of PII and PII stewardship.

The first sub-research question aims to define and conceptualise PII. The goal is to conceptualise PII, capturing its key characteristics and differences from an information infrastructure, the complexity involved in designing PII, and its entire design life cycle. To address this, we conduct a literature review of the concept and related terminology.

We began with a broad explanation of the concept of information infrastructures, based on existing theory. We recognise that a variety of terms are used to refer to PII within the scientific community. However, the aim of this study is not to establish a standardised terminology, but rather to clearly define the specific object of inquiry. Consequently, we adopt the term '*information infrastructures*' and specify the particular meaning given to public information infrastructures in this context. Here we identify that a PII exist out of three components: data, technical elements, and governance structures, as shown in

Figure 2-2. Data can be in one or more formats exchanged between actors in the PII. The technical component of a PII includes systems interconnected to exchange data, standards such as exchange or security protocols, and services provided for data exchange, such as validation or transformation. Finally,

governance can be represented by a single governance structure or by several governance structures. Further details on this topic are provided in Chapter 5.3.2.

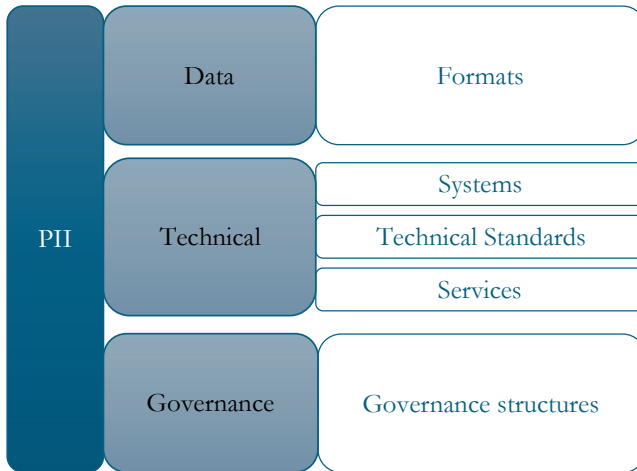


Figure 2-2. The components of a PII

We draw on existing academic literature in fields such as information infrastructure, digital government, and public administration, among others. In developing this conceptual framework, we used a snowballing approach, which enabled us to identify relevant literature through references cited in the initial set of sources. This process highlighted several influential authors whose work has significantly shaped the field. Their contributions have been incorporated into our analysis, thereby enriching and refining our understanding of PII. This approach provides a flexible, comprehensive theoretical understanding of PII and its design.

We applied the same process to the concept of PII stewardship. However, there is limited research on stewardship at the PII level. Based on the characteristics needed for maintaining, designing and reshaping a PII, concepts such as governance and data stewardship were considered relevant but insufficient to fully define this emerging form of steering. Governance is commonly used as a broad conceptual framework, but it does not explicitly address safeguarding the quality of PII. Conversely, data stewardship offers a more operational perspective on ensuring quality dimensions, but it lacks the high-level, overarching view required to cover the entire scope of PII. Based on our review of the existing literature, we develop a tentative definition of PII stewardship.

While there is some literature on quality in Information Systems, we lack an understanding of quality in a PII. Based on the previously identified PII components, data, technical and governance components, a PII steward should aim for high quality in each of these components. Therefore, we suggest that PII quality consists of three dimensions: data quality, technical quality, and governance quality. Each of these quality

dimensions has a specific set of measures. Figure 2-3 presents the relation between PII quality, the three dimensions and their measures.

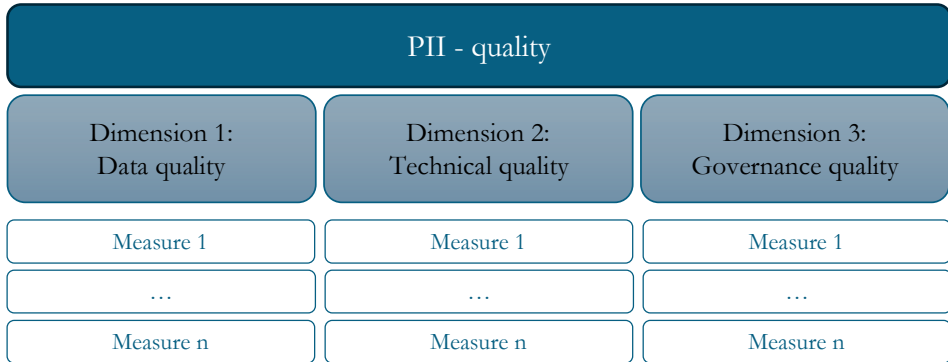


Figure 2-3. Relation between PII quality, dimensions and measures

To establish the measures of data quality and technical quality in a PII, we drew on the lists of data quality and system quality measures compiled by Delone and McLean (1992). There is no overview yet available of the measures of governance quality. Therefore, we conducted a literature review to identify these measures from various authors. This resulted in a framework of three categories (data-, technical-, and governance-quality), with each category including multiple measures.

We have developed a new framework that integrates quality measures identified in existing literature. While this approach provides a structured foundation, it also represents an emerging and relatively new field. Therefore, it is crucial to validate this framework and ensure the list of variables is both comprehensive and relevant. We therefore employ an empirical validation of the quality framework as presented in the following sub-section.

2.3.2 EMPIRICAL VALIDATION OF QUALITY FRAMEWORK

Since it is not immediately clear which measures should be incorporated into the PII framework, we employ three PII cases to verify the literature-based measures and identify potential new quality measures. Based on the criteria in Table 1, we identified three matching cases. These three cases include (1) Customs declarations through Single Window, (2) Tax reporting through Standard Business Reporting, and (3) eInvoicing messages through Peppol-network and Digipoort. The cases are presented in Chapter 4.1.

Table 1. Case selection

#	Criteria	Definition
1	X2G	Data exchange with a governmental organisation
2	Data-driven government	The data demanded by government agencies is essential to carrying out their primary functions and fulfilling their public duties.
3	Multiple actor groups	At least three groups of actors are involved: government agencies, data providers (businesses), and intermediaries.
4	Volume	More than a million messages are exchanged each year.
5	Openness	Any organisation, public or private, can join the PII under a minimal set of entry conditions (i.e., adopt one or more standards).
6	Variety	There are a variety of channels (e.g., system-to-system or portal-based) and components that can be used for data exchange.

The cases are studied by analysing documents and conducting expert interviews. First, we will search for publicly available documentation for each case. For Standard Business Reporting and eInvoicing, many documents are available online, whereas most documentation for the Single Window is provided by the interview respondents. The documents used include descriptions of governance, data exchange, standards, and the systems used. Each document was analysed for quality measures.

Secondly, we conducted interviews with nine respondents from the three cases. The quality framework developed in step 2.3.1 is utilised to create a semi-structured interview protocol. We carried out a total of nine semi-structured interviews (three for each case) to discuss quality measures of a PII.

The nine interview respondents were chosen based on three criteria. First, they must have over five years of experience in the PII, ensuring they possess in-depth knowledge of the subject. The second criterion was that respondents held a tactical role (e.g., programme managers, project managers), ensuring they had expertise in PII governance and technology levels. The third and final criterion was that all respondents represented the government in the PII. This allowed us to capture the government’s perspective on PII stewardship. Based on these criteria, we contacted nine respondents. All nine were willing to participate in an interview. An overview of the respondents is presented in Table 2.

Table 2. Respondents for identifying PII quality measures

#	Role	Area of expertise	Case
1.1	Public intermediate organisation, standardisation office	Technical & Governance	eInvoicing
1.2	Public intermediate organisation, supervision	Technical & Governance	eInvoicing
1.3	Public intermediate organisation	Technical	eInvoicing
1.4	Public intermediate organisation	Governance	SBR – Tax Administration
1.5	Public intermediate organisation	Governance and standards	SBR – Tax Administration
1.6	Public intermediate organisation	Technical	SBR – Tax Administration
1.7	Policy maker	Governance and holistic perspective	Customs declaration through Single Window
1.8	Receiving organisation	Governance	Customs declaration through Single Window
1.9	Public intermediate organisation	Technical	Customs declaration through Single Window

During the interviews, the literature framework and the findings from the document research were validated. The interview format can be retrieved online from the TU Delft repository under the DOI: [10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45](https://doi.org/10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45).

We asked respondents to reflect on their own experiences during interviews lasting between sixty (60) and ninety (90) minutes. The interviews were audiotaped, transcribed, and sent back to the respondents for validation. Four out of the nine respondents provided further details in response to the validation email. The interview transcripts were analysed using Atlas Ti. Our coding framework was initially informed by three predefined themes derived from earlier research: data quality, governance quality, and technical quality of a PII. Within these themes, several quality measures from the literature had already been identified and incorporated into a preliminary codebook. For example, in data quality, there are measures such as consistency, accuracy and timeliness. This codebook provided structure and consistency during the initial coding phase. However, to ensure we captured the quality measures of a PII, we also actively looked for new and alternative quality measures that were not previously documented. The final codebook can be retrieved online from the TU Delft repository under the DOI: [10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45](https://doi.org/10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45). Parts of this research have previously been published in a paper by van Donge et al. (2022). The results of this part of the research are presented in Chapter 3.3.

2.4 RESEARCH PHASE 2: *DESIGN VARIABLES*

The second phase, of which the results are presented in Chapter 5, gains insight into the different design variables in PII.

RQ 3

What are the relevant design variables of public information infrastructures?

Systematic Literature research and empirical research through three cases, including document analysis and interviews

Outcome: Framework for (relevant) design choices in PII

As shown in Figure 2-1 at the beginning of this chapter, to answer this sub-research question, we conducted both a systematic literature review (SLR) towards design choices and an empirical case study. We began with the SLR to identify potential design choices. Based on this list, we conducted interviews to validate and develop a comprehensive framework of PII design choices and variables.

2.4.1 SYSTEMATIC LITERATURE REVIEW TOWARDS DESIGN CHOICES

The aim of the systematic literature review is to identify relevant design variables discussed in the literature. We examined previous research to compile a general list of design variables that aid in understanding how to design a PII. During the earlier literature search, it became clear that no literature has been written on PII design variables as a concept. Therefore, it is necessary to broaden the scope towards information sharing with governmental organisations and analyse the literature on potential design choices or design variables.

We use the Systematic Literature Review (SLR) method, a more structured approach for identifying, evaluating, and synthesising existing research on a clearly defined topic (Page et al., 2021). This method allows other researchers to replicate the review. We follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure that the review is replicable and meets scientific standards. As with any research, it is essential for an SLR to have high validity, reliability, and repeatability (Xiao & Watson, 2019). The PRISMA framework supports these criteria by providing detailed protocols. For instance, it emphasises the importance of defining a clear search strategy in advance, specifying the information sources to be used, and outlining inclusion and exclusion criteria for selecting studies (Page et al., 2021). The PRISMA methodology is illustrated in Figure 2-4.

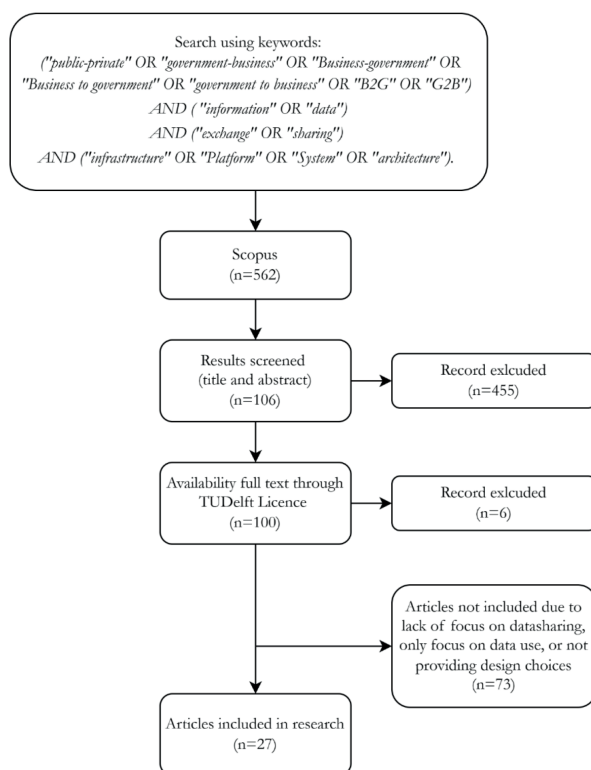


Figure 2-4. PRISMA systematic literature review

We systematically searched the electronic database ‘Scopus’, as this is a highly valued database for the Dutch scientific system.

We did not use the search terms ‘design choice’ or ‘design variable’ as this was too specific and only resulted in six papers. All of these six papers are also included in the search string without using the terms ‘design variable’ or ‘design choice’. Instead, we searched for ‘infrastructure’, ‘platforms’, ‘system’, or

'architectures', which helped us find research on PII where we could define the design variables or choices ourselves. To maximise the number of eligible studies, we used the search string: ("public-private" OR "government-business" OR "Business-government" OR "Business to government" OR "government to business" OR "B2G" OR "G2B") AND ("information" OR "data") AND ("exchange" OR "sharing") AND ("infrastructure" OR "Platform" OR "System" OR "architecture").

The following criteria are used in the literature search:

Field: Articles should study information sharing with the public sector, as the objective of this study is to promote high-quality information exchange with governmental organisations.

Topic: Studies included needed to provide details on information infrastructures. Research focusing on the use of information is not included in this review. Design variables or choices needed to be mentioned in the paper. Papers on barriers to information infrastructures, for example, did not offer insights into the choices practitioners have when designing a PII. Papers with actual design variables mentioned (by explaining the case), and papers discussing a design variable, are included in this review.

Language: This review includes only articles written in English.

Publication types: Only peer-reviewed books, articles, conference papers, and chapters from books are included in this review.

Year of publication: The literature search covered the period from 2000 to November 2023.

The study selection process is illustrated in the PRISMA flow chart in Figure 2-4. The initial literature search in Scopus resulted in 562 records. The criteria for including a study are the focus on data sharing and providing insights into possible design choices in a PII for the data, technical and governance components. As stated in Chapter 1.3, a PII design variable should meet the following criteria:

Scope: The variable relates to the design of the PII, including its data, technical, or governance components.

Configurability: The variable can be deliberately chosen or configured by relevant stakeholders.

Impact: The variable influences the infrastructure's characteristics or performance (e.g., safety, interoperability, or efficiency).

Measurability: The variable can be measured or assessed systematically.

Discrete options: The variable can be expressed through clearly distinguishable design options or configurations.

Relative independence: The variable is to a large extent independent of other design variables.

In the first stage, the articles were assessed through a quick scan of titles and abstracts. This process resulted in 106 relevant articles remaining. In the second step, we downloaded all the articles to read their full texts. Of these, six sources were not available as full texts; these articles are presented in the second table in Appendix A. They were excluded before the full eligibility assessment. The available full texts were evaluated based on topic and field criteria. Articles that focused on information sharing, mentioned design variables or choices, or involved the public sector were excluded from the research. This resulted in twenty-seven articles being considered relevant to this study. The first table in Appendix A presents the literature used in the review.

2.4.2 EMPIRICAL RESEARCH TO VALIDATE AND COMPLEMENT THE DESIGN FRAMEWORK

The main objective in this phase is to answer the sub-research question: **What are the relevant design variables of a public information infrastructure?** To answer this question, we started by building a framework for the design variables of PII. Through an empirical case study of the three cases presented in Chapter 4.1, we identify additional variables and validate those identified in the literature for designing a PII. The results are presented in Chapter 5.

In this second empirical study, we used the same three cases, Customs declaration through Single Window, Tax declaration through SBR and eInvoicing to the central government, as in phase one, to maintain continuity and build upon existing insights. This means we used the same selection criteria as outlined in phase 1. Reusing previous cases allows us to gain a deeper understanding of the cases and the design choices made by the actors in the PII. Additionally, it provided insights into how these choices relate to the available options and different dimensions of quality. Although this study does not track developments over time, engaging with familiar cases increased the focus and enabled more in-depth exploration. Relying on similar cases might limit the generalisability of the findings. To address this, we will analyse over twenty cases in the third phase to explore this further.

We started by analysing the available documentation. Through content analysis, we examined the documents for potential design variables and contextual factors. Next, using semi-structured interviews, we validated, expanded, and refined the literature and documentation related to design choices. To achieve the appropriate level of abstraction for each variable, this research employed semi-structured interviews. The interview protocol can be retrieved online from the TU Delft repository under the DOI: [10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45](https://doi.org/10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45).

Because we use the same cases, we also interviewed some of the same respondents, as they were deemed experts in this case. During the first phase, these respondents primarily focused on quality aspects in PII. However, it became apparent during the interviews that they also had in-depth knowledge of PII design choices. To obtain a more complete and balanced picture of the PII, additional respondents were selected. Additional criteria for respondents are that they must have over five years of experience in their PII case, ensuring sufficient expertise. Secondly, respondents need to hold a role at the tactical level, such as a programme or project manager. The third and final criterion is that all respondents represent the government within the PII. A conscious effort was made to ensure diversity in roles and areas of expertise. Through snowball sampling, we added eight new respondents, resulting in seventeen interviews with a total of eighteen respondents. Table 3 presents the respondents interviewed in the second research phase.

Table 3. Overview of respondents interviewed in research phase 2

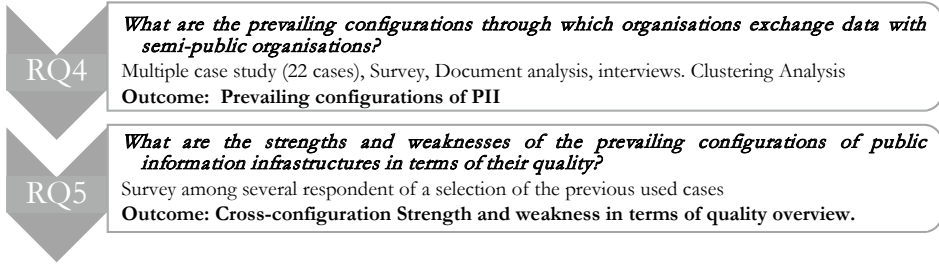
Reference number	Rol	Area of expertise	Case
2.1	Public intermediate organisation	Technical expert & Governance expert	eInvoice
2.2	Policy maker	Governance expert	eInvoice
2.3	Supporting organisation	User support expert	eInvoice
2.4	Supervising authority & public intermediate organisation	Technical expert & Governance expert	eInvoice
2.5	Public intermediate organisation	Technical expert & Governance expert	eInvoice
2.6	Public intermediate organisation	Technical expert	eInvoice
2.7	Public intermediate organisation	Technical expert	SBR
2.8	Public intermediate organisation	Governance expert	SBR
2.9	Public intermediate organisation	Governance expert	SBR
2.10	Public intermediate organisation	Governance expert	SBR
2.11	Public intermediate organisation	Technical expert	SBR
2.12	Receiving organisation/policy maker	Governance expert	SBR
2.13	Public intermediate organisation	Technical expert	Customs declaration
2.14	Public intermediate organisation	Governance expert	Customs declaration
2.15	Policy maker	Governance expert	Customs declaration
2.16	Public intermediate organisation	Technical expert	Customs declaration
2.17	Public intermediate organisation	Technical expert	Customs declaration

The interviews were audiotaped, transcribed, and sent to the respondent for validation. Three out of the seventeen respondents provided additional details in response to the validation email. The interview transcripts were analysed using Atlas Ti. We conducted a thematic coding process based on a hybrid approach that combined both deductive and inductive strategies. Our coding framework was initially informed by three predefined themes derived from earlier research: data, governance, and technical implementation of a PII. Within these themes, several design choice variables from the literature had already been identified and incorporated into a preliminary codebook. This codebook provided structure and consistency during the initial coding phase. However, to ensure we captured the design options in a PII, we also actively looked for new design variables and alternative design options that were not previously documented. Additionally, we inductively identified an initial set of contextual variables, which will be validated in the third research phase. This inductive element enabled us to extend the existing framework and reflect new insights emerging from the empirical data. The final codebook can be retrieved online from the TU Delft repository under the DOI: [10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45](https://doi.org/10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45). Parts of this research have previously been published in a paper by van Donge et al. (2024). The outcome of this research is presented in Chapter 5.

2.5 RESEARCH PHASE 3: CONFIGURATIONS

A configuration of a PII consists of choices made regarding the data exchanged within the PII, the technical components facilitating the exchange and additional services, and the governance structures involved in coordinating, managing, and overseeing (parts of) the infrastructure. In the third phase, of which the results

are presented in Chapter 6, this research uses a multiple-case study design for cross-case analysis to explore how different cases are configured in terms of data, technical and governance design choices. We expect cases to have made different design choices, leading to varied configurations. However, we anticipate that some choices will be more common in certain instances, forming prevailing configurations. This phase delves deeper into the configuration of design choices and assesses their strengths and weaknesses in terms of quality. It aims to answer the following sub-research questions.



As shown in Figure 2-1 at the beginning of this chapter, this phase uses surveys, document analysis, and semi-structured interviews, to come to the configuration of the PII cases, and through configurational analysis find empirically derived prevailing configurations of PII design choice.

2.5.1 PREVAILING CONFIGURATIONS

To identify prevailing configurations, we require sufficient data to reveal common patterns across the cases. For this study, a purposive case selection was used to gain in-depth insight into the design clusters in PII. We selected multiple cases to allow a deeper understanding of the various configurations of information infrastructures. Table 4 presents the criteria for case selection. We identified multiple cases and subsequently made a selection based on predefined criteria. If, after an interview, survey, or document analysis, a case appeared not to meet these criteria, it was excluded from the study. Additionally, we focused on identifying PIIs that were sufficiently similar according to the criteria presented in Table 4, while still differing sufficiently in their technical components, stakeholders, and data objectives to allow meaningful comparison. The selected cases are presented in Chapter 3.2, and a more in-depth description is provided in Appendix B. Using twenty-two cases contributes to comparability and contextual richness, allowing for hierarchical clustering and cross-case analysis.

Table 4. Case selection criteria for the multiple-case study

#	Criteria	Definition
1	X2G context	The case involves data sharing with a public organisation or an entity performing a public function.
2	Multiple stakeholders	The presence of multiple stakeholders, ensuring a diversity of interactions in the II
3	Available expert	The availability of at least one domain expert with strategic insight into the PII;
4	Country	The case is situated within the Netherlands or serves a purpose directly relevant to the Dutch context.
5	Established PII	The PII is implemented and (to some extent) used in practice. It should be in the life cycle phase of establish, expand, stagnation, transition, and drift. This means PII's in the design phase are not taken into account.

For each case, we selected at least one respondent to be interviewed and to complete the survey. The selection criteria for respondents are similar to those used in phases one and two: a minimum of five years' experience in the domain, a tactical or strategic role ensuring they have an overview of the PII, and representation from semi-public organisations. Based on these criteria, we conducted interviews with twenty-five respondents across twenty-two cases. In the cases of CBS, SBR-education, and Land Registry, interviews were conducted with two respondents simultaneously, at their request. Interviewing two respondents can generate richer data, as they can complement or challenge each other's perspectives. However, they may also influence each other's answers, leading to more socially desirable responses. These drawbacks were mitigated by ensuring both respondents had equal opportunities to speak. Additionally, the interviews with respondents were conducted after the surveys, which limited the sensitivity of the questions. All respondents completed the survey for their case. An overview of the respondents is presented in Table 5.

Table 5. Respondent interviews and survey phase 3

Reference number	Casus	function	Type of organisation
3.1	Pension funds	Strategic advisor	Receiver
3.2	BRP issue reporting	Business consultant	Public intermediate
3.3	eInvoicing	Architect e-procurement Standards	Public intermediate
3.4	Bank supervision	Innovation advisor	Receiver
3.5	SBR-education	Advisor technology	Policy maker & receiver
3.6	SBR-education	Advisor governance	Policy maker & receiver
3.7	eCodex	Program-manager	Public intermediate
3.8	Edu-V	Program-manager	Supporting organisation
3.9	Import declaration	Advisor innovation	Receiver
3.10	WMO	Coordinator infrastructure	Policy maker
3.11	Child authority	Enterprise architect	Policymaker
3.12	Vaccination	Enterprise architect	Supporting organisation & intermediate
3.13	Land registry	Board member	Receiver
3.14	Land registry	Architect	Receiver
3.15	Kik-V	Advisor	Supporting organisation
3.16	National housing progress	Advisor data	Representative receivers
3.17	Wage returns statistics	Strategic manager	Receiver
3.18	Wage returns statistics	Account manager	Receiver
3.19	Migration	Business architect	Receiver & policy maker
3.20	CareCodex	Executive director	Supporting organisation
3.21	OOTS	Architect	Public intermediate
3.22	Single Window	Governance advisor	Receiver & policy maker
3.23	VUM	Sr. Advisor	Policy
3.24	WOCO	Business consultant	Public intermediate
3.25	Notification property value	Architect	Public intermediate

To gain deeper insight into the cases before conducting an interview, we asked respondents to complete a survey in advance and to provide additional documentation of their case. In the survey research, we will ask participants how the information infrastructure is configured. The survey questions can be retrieved online from the TU Delft repository under the DOI: [10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45](https://doi.org/10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45).

To analyse the configuration of a PII, we reviewed available online documents and those provided by the respondents. As previously mentioned, we view an information infrastructure as consisting of data, technical, and governance components. Therefore, we reviewed documentation covering these three aspects. If there is a discrepancy between the survey data and the documented data, the interview offers an opportunity to clarify the matter.

Based on the survey and document research, we can have a more in-depth discussion on the configurations during the semi-structured interviews. The preparation shows where more insights were needed, and the focus should be during the interviews. For the interviews, we use a semi-structured interview

protocol. This protocol can be retrieved online from the TU Delft repository under the DOI: [10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45](https://doi.org/10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45).

The analysis of the data starts with searching for patterns and insights in the configuration. The documentation and interviews will be analysed using qualitative coding by multiple researchers to enhance the trustworthiness and reliability of this research phase.

All interviews were audiotaped, transcribed, and sent to the respondents for validation. Three of the twenty-five respondents have provided more details in response to the validation email.

The interview transcripts are analysed using Atlas.ti. For this third phase of the research, we applied a deductive coding approach, guided by the three predefined overarching components of a PII: Data, technology, and governance. We included the predefined set of design variables and context variables from phase two as codes in our codebook. Although our coding framework is primarily driven by prior research, we intentionally left space for inductive insights. This allows us to identify other design choices that were not captured by earlier research. The final codebook can be retrieved online from the TU Delft repository under the DOI: 10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45. By combining the survey data with the interview findings, we constructed an overview of the design choices present in each case.

The presence of multiple configurations of the design choices across the 22 PII cases enables us to cluster and identify patterns in these configurations and detect prevailing configuration types. Hierarchical clustering was selected as the primary clustering method for this study. Alternative approaches, such as centroid-based clustering, fuzzy set clustering, and latent class clustering, were also considered. However, the absence of a clear outcome variable, the nature of the variables in the dataset, and the limited number of cases made these approaches less suitable. The objective of this study is to explore patterns and relationships among design variables across cases rather than to explain or predict a dependent variable. Hierarchical clustering allows for a step-by-step examination of similarities between cases and offers a clear visual representation of the clustering process through a dendrogram. It is particularly suitable for exploratory research where the goal is to identify prevailing configurations among cases based on multiple categorical or mixed variables (Kaufman & Rousseeuw, 2005).

A disadvantage of hierarchical clustering is that its results are influenced by the scale and number of variables. Adding more variables can alter the clustering structure because each variable influences the distance calculation. When there are too many variables, noise may overpower the true signals, making it difficult to identify clear and meaningful clusters. Conversely, with too few variables, important patterns or dimensions might be missed, causing relevant clusters to remain hidden. Therefore, we decided to identify sub-configurations for each of the PII components separately. We start by identifying the sub-configurations within the data, governance, and technical components. We cluster the data based on the design choices made for these specific components. Data design choices are used to identify sub-configurations within data components, technical design choices are used to find sub-configurations within technical components, and governance design choices are utilised to determine the sub-configuration of governance components. Using these sub-configurations, we created an overview of the case and compared the configurations that

exist out of sub-configurations. Here, we compared the cases for recurring patterns of sub-configurations that lead to the prevailing configurations.

For each of the sub-configurations, we performed a clustering analysis separately. This analysis provides deeper insight into each of the components of a PII. We used hierarchical clustering in Python. The Python code used can be retrieved online from the TU Delft repository under the DOI: [10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45](https://doi.org/10.4121/50df56a6-4d14-459d-a5fb-cb96f1bf0d45).

To establish the sub-configurations, we use a hierarchical clustering method. Clustering techniques often require an estimate of the number of clusters in advance, whereas hierarchical clustering is a more flexible technique for clustering the data objects (Aggarwal, 2018). Moreover, in hierarchical clustering, the data are not confined to one specific cluster that stands out from others but instead have a cumulative structure. They can be arranged to range from a single cluster containing all cases to each cluster consisting of a single case (Everitt et al., 2011). This type of clustering provides an overview of the hierarchical clusters in a dendrogram, which illustrates the convergence or divergence of variables.

An important aspect of cluster analysis is deciding when a group of observations can be considered a meaningful cluster. Clustering is essentially an exploratory method where the number of clusters is not predetermined but arises from the data's underlying structure. However, a thorough interpretation of the results requires clear criteria for assessing the robustness and relevance of clusters.

Hierarchical clustering can be either agglomerative or divisive. Agglomerative, or bottom-up, means that clustering begins with a single point and merges points based on their (dis)similarities (Aggarwal, 2018; Everitt et al., 2011; Kaufman & Rousseeuw, 2005). Divisive clustering, or top-down clustering, starts with all data points in a single cluster and repeatedly splits them into smaller groups. Divisive clustering can be used when prior knowledge of how to split the dataset is available, which is useful when lower-level details are less important (Tan et al., 2006). In our research, we chose to cluster without any subjective judgment, therefore opting for agglomerative clustering. Additionally, agglomerative methods are suitable for smaller datasets. In these techniques, a dissimilarity matrix is used to represent the dissimilarities among data observations. Depending on the clustering method, similar objects are clustered. This process repeats until all variables are in a cluster (Aggarwal, 2018).

In Agglomerative clustering, single linkage and complete linkage are the most common techniques. Many clustering methods require Euclidean-type variables. Euclidean variables are numerical variables that can be represented as points in a geometric space (Kaufman & Rousseeuw, 2005). However, the variables within this research are nominal. Variables can be binary (two states, often coded as 0 or 1) or have multiple states (Kaufman & Rousseeuw, 2005). An example of this type of variable is the organisation of the technical infrastructure. Here, we observe several states, including: (1) central system, (2) semi-central, (3) decentral, (4) distributed, or (5) hybrid. Both single linkage and complete linkage are appropriate for this type of nominal data. It is essential to select the correct measure for cluster analysis, as different measures of (dis)similarity can produce distinct clusters (Everitt et al., 2011).

In single linkage, *“the similarity of two clusters is the similarity between their most similar (nearest neighbour) members. This method intuitively gives more importance to the regions where clusters are closest, neglecting the overall structure of the*

cluster.” (Aggarwal, 2018, Chapter 4.3.1.1). However, in our analysis, we want to account for the overall structure of the clusters and identify distinct cluster types. “*Complete linkage (or furthest neighbour) is opposite to single linkage, in the sense that distance between groups is now defined as that of the most distant pair of individuals.*” (Everitt et al., 2011, p.62). In complete linkage, the distance between two clusters is defined as the greatest distance between a pair of observations, one from each cluster. “*This is equivalent to choosing the cluster pair whose merge has the smallest diameter. As this method takes the cluster structure into consideration, it is nonlocal in behaviour and generally obtains compact-shaped clusters.*” (Aggarwal, 2018, Chapter 4.3.1.1). This method is more robust to noise and outliers than, for example, single linkage. It typically results in compact, well-defined clusters (Kaufman & Rousseeuw, 2005). Additionally, a method such as Ward would not be suitable because it assumes that the underlying distance measure is a Euclidean metric (Everitt et al., 2011). Research shows that the complete linkage method tends to create meaningful clusters (Shao, 2007). Therefore, the complete linkage method was chosen as the clustering criterion. Complete linkage can be calculated as shown by (Scitovski et al., 2021):

$$D_{max}(A, B) = \max_{a \in A, b \in B} d(a, b)$$

To create a dissimilarity (or distance) matrix, we need to have a dissimilarity coefficient. The dissimilarities are non-negative numbers $d(i, j)$. If two cases (i and j) are more similar, the number is relatively low, whereas if they differ more, the number becomes larger. This is represented in the distance matrix. The hierarchical cluster analysis was performed using the Jaccard distance for binary datasets. The Gowers method was selected exclusively for Governance variables because it contains ordinal variables. In our cases, we coded the variables asymmetrically. For asymmetric data, the following applies: “*By convention, we shall always code the most important outcome (which is typically the rarest one) by 1, and the other by 0.*” (Kaufman & Rousseeuw, 2005, p.26). We code each outcome as the presence or absence of a specific characteristic. Then we examined the cases to identify similarities in the presence of the variables.

The most well-known distance coefficient for asymmetric binary data is the Jaccard coefficient (Kaufman & Rousseeuw, 2005). The dataset used consists of binary variables indicating the presence (1) or absence (0) of specific characteristics per observation. In such cases, the Jaccard distance is a suitable measure because it is explicitly designed for binary data. The Jaccard distance focuses on features that are actually present in at least one of the observations (Kaufman & Rousseeuw, 2005). Consequently, only positive matches (the shared presence of relevant features) are used to indicate similarity. This is especially valuable in contexts where the presence of a feature is informative (Kaufman & Rousseeuw, 2005).

In addition to using the Jaccard distance for binary data, this study also applied the Gower distance to the governance variables dataset. The dataset contains both categorical and binary variables. The Gower distance allows for distance calculations in datasets comprising different types of variables (Kaufman & Rousseeuw, 2005). The application of Gower distance is critical in situations where reducing the dataset to binary variables would result in a loss of vital information (Everitt et al., 2011). In the dataset in question,

both nominal categories and binary indicators were present. The Gower distance enabled the inclusion of all relevant variables simultaneously in the distance calculation in a methodologically sound manner, without requiring unnatural data transformations.

From this analysis, we generated distance matrices and corresponding dendrograms, which are presented in Chapters 6.1, 6.2, and 6.3. In the research, the final number of clusters was determined based on a combination of dendrogram analysis, visual inspection of the cluster matrix, and qualitative assessment of cluster relevance. This resulted in a clustering of substantially related groups, where observations within a cluster are strongly similar to each other based on the variables used, while clearly distinguishable from observations in other clusters. This approach enhances the validity of the clustering and facilitates a meaningful interpretation within the context of PII.

2.5.2 EXPLORING STRENGTHS AND WEAKNESSES

Within this study, in phase one: *Definitions*, we established a PII quality framework. However, it became apparent that a more extensive study was necessary to examine the PII quality in detail. We opted for a more straightforward approach to measure configuration quality, as we still aim to say something meaningful about quality. Consequently, we identified the strengths and weaknesses of the PII in terms of data, technical, and governance quality.

From each configuration, two cases were selected for further analysis. The primary respondents in these cases were asked to share the survey with three colleagues in their field. These secondary participants were invited to share their perspectives on the quality of the PII, specifically regarding strengths, weaknesses, opportunities, and threats/risks. Respondents were asked to complete an online survey that included a question assessing the case across three quality dimensions: data quality, technical quality, and governance quality.

Of the thirty-six surveys sent out, we received twenty responses, thirteen of which were completed in full. Table 6 provides an overview of the various respondents. No responses to the surveys were received from the technical-driven configuration.

Table 6. Respondents - Survey PII quality

#	PII Configuration	Public or private organisation	Function	Casus
4.1	Mixed collaboration	Public	Policy maker	WMO
4.2	Mixed collaboration	Private	Providing organisation	SBR-Education
4.3	Mixed collaboration	Public	Receiving organisation	SBR-Education
4.4	Mixed collaboration	Public	Receiving organisation	SBR-Education
4.5	Mixed collaboration	Public	Developer standards	eInvoicing
4.6	Mixed collaboration	Public	Policy maker	eInvoicing
4.7	EU-regulation	Public	Receiving organisation	Banking supervision
4.8	EU-regulation	Public	Intermediary/service provider	Banking supervision
4.9	Central data reuse	Public	Receiving organisation	Wage returns statistics
4.10	Central data reuse	Public	Receiving organisation	Pension funds
4.11	Public collaboration	Public	Developer standards	Notification property value
4.12	Distributed	Public	Implementation support	VUM
4.13	Distributed	Public	Project-lead	VUM

Furthermore, the interviews conducted during earlier research phases also offered insights into strengths and weaknesses. These findings were incorporated as input for this part of the study and systematically coded during analysis. They are included in the development of potential quality characteristics for each configuration. Specifically, for the technology-driven configuration, only items derived from interviews were used to provide insights into its possible quality strengths and weaknesses.

2.6 RESEARCH PHASE 4: STEWARDSHIP

In the final research phase, of which the results are presented in Chapter 7, we elaborate on the empirically derived classifications of PII stewardship and answer the following sub-research question:

RQ 6

What classification of public information infrastructure stewardship can be distilled from the findings of this study?

Combining research step 1, the conceptual synthesis with data from research step 3 with secondary data synthesis

Outcome: deepening of the concept PII stewardship with a PII stewardship classification.

This study adopted a qualitative, contemplative approach, merging a conceptual synthesis from research phase one: *Definitions*, with a secondary data analysis of the cases used in research phase three: *Configurations*. The aim was to deepen the theoretical understanding of PII stewardship by integrating insights from the literature with empirical findings from this research to answer sub-research question six.

In phase one: *Definitions*, a conceptual analysis was conducted to clarify the concepts of PII and PII stewardship, drawing on relevant theoretical sources. In research phase three: *Configurations*, we analyse multiple cases to establish configurations of PII. Using these cases and their configurations, we can analyse the stewardship roles associated with them. This secondary data enables us to enrich our conceptual analysis of PII stewardship and to empirically derive with a classification of PII stewardship.

2.7 CHAPTER CONCLUSION

This chapter describes the mixed-methods research approach used in this study, which combines quantitative and qualitative methods. This enables examination of PII and PII stewardship from multiple perspectives, therefore enhancing the depth, validity, and reliability of the results.

The research was carried out in four structured phases, each addressing one or more of the six sub-research questions. In the first phase, *Definitions*, the concepts of PII, PII stewardship, and PII quality are examined. The second phase, *Design variables*, involves conducting a systematic literature review to inform the design choices for a PII. These choices are then validated through a multiple-case study involving three cases, which used semi-structured interviews and document analysis. The third phase, *Configurations*, focuses on exploring the existing configurations and their strengths and weaknesses through a multiple-case study, enhanced by document analysis, survey research, and semi-structured interviews. In the fourth phase, *Stewardship*, we synthesise to systematically consolidate insights from the literature review, case studies, document analysis, survey results, and semi-structured interviews. This provides an understanding of the categories of PII stewardship in practice.

This phased approach enables the systematic development, testing, and refinement of insights as the research progresses. By combining these methods and adopting an iterative research design, a strong and comprehensive body of evidence has been built. The following chapters present the outcome of phase one, *Definitions*, providing insight into PII, PII stewardship, and PII quality

Chapter 3

THEORETICAL BACKGROUND

In this chapter, we present the findings of the first research phase: Definitions. We begin by examining the theoretical foundations for understanding the design and management of information infrastructures within the context of data exchange with (semi-)public organisations. These theoretical foundations are supported by empirical research, providing an overview of information infrastructures. Guided by two central sub-research questions, we developed a conceptual framework based on both the literature and empirical evidence.

Section 3.1 addresses sub-research question one: **What is a public information infrastructure?** We focus on the different characteristics of an information infrastructure and on what constitutes PII, specifically developed for data sharing with (semi-)public organisations. From the literature, we identified five characteristics of an information infrastructure that can also be used to characterise a PII. For each of these characteristics, we explore additional challenges for PII. Due to the increased complexity and configurational dynamics described in the introduction, PII stewardship represents the authorities for designing, maintaining, and adapting the PII configuration, and ensuring alignment and focus on quality within a PII. We propose a PII quality framework in Chapter 3.3, as PII stewards should strive for in the design of a PII. By proposing a tentative definition of PII stewardship, section 3.2 addresses sub-research question three: **How can public information infrastructure stewardship be defined?**

Together, these sections aim to deepen our understanding of the different dimensions of public information infrastructures and PII stewardship.

3.1 PUBLIC INFORMATION INFRASTRUCTURES

A public information infrastructure (PII) is an information infrastructure explicitly developed for sharing information with the public sector.

Since limited research has been conducted on PII, we first examine the general characteristics of information infrastructures. Information infrastructure refers to (digital) facilities and services necessary for performing human activities and can be compared to ‘enabling resources in network form’ (Bowker et al., 2010, p.2). Although they are part of everyday life, infrastructures often remain invisible or taken for granted. While there is extensive research on information infrastructures within the field of information systems, the definition of an information infrastructure remains a topic of ongoing discussion. An essential characteristic of (public) information infrastructure is that it comprises multiple technical components (Hanseth & Lyytinen, 2010). Multiple systems are interconnected, and services such as data validation can be implemented to enhance efficiency. Due to the connections between various systems, interoperability of data through formats and adherence to standards are becoming increasingly important. However, they are not merely technical systems. Because they connect multiple systems, they also serve as links between various users. Therefore, the third characteristic of a PII is the involvement of multiple stakeholders. In addition to the expected end users of the PII, the organisations involved in data sending and receiving, there is a trend of intermediary organisations providing systems and services within the PII. The functionalities

needed for the information infrastructure may already be developed by other organisations, reducing the need to build from scratch (Janssen, Chun, & Gil-Garcia, 2009).

Due to the dynamics between technical systems and social interactions, an information infrastructure is considered a socio-technical system. *“Most infrastructures are in part engineered, after which they evolve.”* (Janssen et al., 2009, p.234). Therefore, a PII has to be researched and maintained as a whole, from both technical and social perspectives. When looking at the development of (public) information infrastructure literature, it is observed that an information infrastructure evolves through adaptation and reorganisation.

In the following subsections, we elaborate on the characteristics of a PII: first, its technical characteristics; then, its standards; its social characteristics; its socio-technical aspects; and, finally, its development.

3.1.1 TECHNICAL SYSTEMS

The primary characteristic of a PII is that it comprises multiple technical components and systems. Although the scope of the infrastructure may vary, it is widely recognised as a chain or network of interconnected systems through which information is exchanged. Some view an information infrastructure as local infrastructures (Shin et al., 2006), while others interpret information infrastructures as global and national networks capable of adapting to different circumstances (Monteiro & Hanseth, 1996). Depending on its scope, infrastructures are relational. An information infrastructure consists of interconnected IT solutions. While previously focused within a single organisation, over time, information infrastructures have evolved into networks involving multiple actors, both public and private. *“As the number of IT devices increases, they spread across more and more application fields (professional as well as private; wireless technologies keep them constantly connected to each other.”* (Pipek et al., 2009, p.448).

Information cannot be exchanged digitally without a technical structure. In the technological components, we distinguish the network of services. *“From a technical viewpoint, designing an information infrastructure involves discovery, implementation, integration, control, and coordination of increasingly heterogeneous IT capabilities”* (Hanseth & Lyytinen, 2010, p.2). Examples of these processes include data preparation, data standardisation (syntax and semantics), data storage, data quality assurance, validation, identification, authentication, authorisation, reconciliation, routing, archiving, and delivery (Bharosa et al., 2018). Both public and private organisations can offer these components as services to assist other entities, thus establishing themselves as intermediaries within the information infrastructure. Information can be exchanged through various interfaces, including services such as portals, software, email, and system-to-system connections. Especially with multiple actors and systems, information can take different forms throughout the entire information infrastructure.

The literature, based on Cordella (2010) and Hanseth and Monteiro (1998), assumes openness of the information infrastructure. This definition of openness implies there are no boundaries to the information infrastructure, and actors can enter and leave an information infrastructure without obstructing borders. This form of participation-type openness, in which participants, such as third parties or end users, can enter

and leave the system, is added to other definitions of openness in research by Ofe and de Reuver (2024). This research shows that openness can also imply editability, interactivity, and distributedness of data (Ofe & de Reuver, 2024). It is important to note that openness is not a default setting in PII. In many cases, we observe that, during data exchange, the editability and interactivity of the data are kept low. For example, in the case of the VAT declaration process, data cannot be altered, is not distributed except to the Tax Administration, and the Tax Administration office performs data format transformations in its secure environment rather than in the information-sharing infrastructure. In a PII, public organisations are often authorities that significantly shape and influence the information infrastructure. Frequently, the public sector determines which standards are used, what the requirements for data supply are, and which actors can and must be involved.

3.1.2 STANDARDS AND FORMATS

The second characteristic is the use of structured data formats and adherence to (technical) standards. *“Information infrastructure and information infrastructure development may, to a large extent, be described as specification and implementation of standards.”* (Hanseth & Monteiro, 1998, p.49). Underlying the services provided in an information infrastructure is a network of systems and hardware that enable data exchange. Henningsson et al. (2011) stated: *“Technically, the construction of an infrastructural system requires the establishment of protocols and standards that enable the system to be used and seamlessly connect with other systems.”*(p.18). To have usable information infrastructures, interoperable standards, definitions, and interfaces are necessary (Filson et al., 2010).

With multiple systems, interoperability is key. *“It are larger and more complex systems, involving large numbers of independent actors as developers as well as users. This fact makes standards a crucial part of information infrastructure.”* (Hanseth & Monteiro, 1998, p.49). To communicate with other parties and understand the data sent, it is essential to standardise interfaces and data formats. The choice of format is often determined by the users of that format (Van de Kaa et al., 2011). In some cases, this is the end user of the information infrastructure. When the end user is not involved in the technique, the designers of intermediary systems take part in using formats. Each user of the format tends to invest in it, making it the most common and widely used format, which ensures interoperability with other systems (Van de Kaa et al., 2011).

It is therefore a dynamic interaction between the developers of the standard and its users within an information infrastructure. Setting a standard requires actors in the information infrastructure to engage and negotiate to reach agreements on these technical components. As the number of parties involved in the exchange process increases, a proliferation of systems and formats is likely, making it more difficult for organisations to exchange information. The article by Van de Kaa et al. (2011) illustrates the factors that influence the outcome of interface format battles and the process of natural selection leading to the survival of certain standards. With new standards emerging and disappearing, the field remains continuously changing. For public organisations, these uncertainties pose significant challenges, as governments are often less flexible in altering their systems, and such changes demand substantial resources. It is difficult to keep up with technological advances, as PII is often inflexible and large-scale. As technology progresses, methods

of exchanging information evolve, and once an information infrastructure component is implemented, further development may follow. Governmental organisations must find strategies to manage these uncertainties. The most effective organisations do not rely on a fixed infrastructure but adapt to changes in their environment (Lewin, 1999).

In an ideal scenario, each government organisation adopts standardised methods for exchanging information, enabling seamless transfer without human intervention and reducing errors. The network externalities, in terms of the benefits of a technology, grow as its usage increases (Van de Kaa et al., 2011). This can only occur if more governmental and private organisations utilise the same technology. However, each government organisation may organise its own systems, services, and standards. Consequently, a variety of standards and specifications exist for supplying organisations, adding to compliance burdens. Standardising existing information flows presents significant challenges, as governmental organisations often have specific requirements. To reduce these uncertainties, the government might opt for a single government-designed protocol or specific PII protocols. Ultimately, agreements may be mandated by regulation or collectively agreed upon, depending on the social structure in the information infrastructure.

3.1.3 SOCIAL CHARACTERISTICS

The third characteristic is the involvement of multiple stakeholders in an information infrastructure. The construction of an information infrastructure requires “*organising and connecting heterogeneous actors with diverging interests in ways that allow for information infrastructure growth and evolution.*” (Hanseth & Lyytinen, 2010, p.2). According to Cordella (2010), infrastructures are phenomena in action. The outcome of the information infrastructure depends on the relationships between the stakeholders. The decision rights within an II are often distributed among the stakeholders participating in the PII, who jointly shape its development. It is a constant interplay between the technology and the actors in the information infrastructure.

Shaping an information infrastructure involves integrating these social and technical dimensions and improving both the technical and social aspects (Cherns, 1976). Individual, organisational, and inter-organisational behaviours are embedded in the technical components of the information infrastructure. Bostrom and Heinen (Bostrom & Heinen, 1977) demonstrated that design changes occur at both the social and technical levels of the information infrastructure. The technical components are shaped through complex social processes. Therefore, the objectives of organisations are not merely to optimise the technical system and align a social system with it, as the technical system often results from social interactions among stakeholders. It requires stakeholders in the information infrastructure to interact and negotiate to establish agreements and standards.

The growing number of stakeholders, and therefore the possible further distribution of decision rights, places greater emphasis on the social components of the information infrastructure. Social structures, roles, and rights are a large part of an information infrastructure. “*Social factors affect the development, adoption, change, operations, and stability of information infrastructure*” (Klievink & Janssen, 2014, p.241). As the number of stakeholders increases, the social components are subject to change. Infrastructures are becoming more

fluid. There is a growing trend towards greater openness regarding PII. Public and private service providers provide services in the PII on behalf of sending or receiving organisations. However, it becomes more challenging to develop infrastructure when it extends beyond organisational boundaries. “*Managing infrastructures transcends the boundaries of centralised, hierarchical control of individual resources*” (Janssen et al., 2009, p.234). The technical infrastructure depends on stakeholders' interactions, negotiations, and decision-making processes. Multiple stakeholders have decision-making rights or seek to influence the infrastructure to suit their needs and goals. The social component involves constant interactions and negotiations between all levels of stakeholders in the infrastructure (Tilson et al., 2010). With more stakeholders, the infrastructure is exposed to a broader range of values, goals, and perceptions of quality. This creates challenges in decision-making.

The number of stakeholders increases in PII when private parties can provide components of the information infrastructure. This may lead to a further distribution of decision-making rights, in which public actors no longer exert direct influence over the choice. The growing number of stakeholders necessitates the formation of new coalitions of partners (Fang, 2002). PII infrastructures involve a variety of heterogeneous stakeholders. These include information providers, information recipients, intermediaries, service providers, policymakers, and regulators.

As the number of parties increases, the system's complexity also rises. Governments rely on the support of other actors to ensure the success of decisions (de Bruijn & ten Heuvelhof, 2008). Due to high stakeholder complexity in public-private collaboration, the role of the social component becomes even more significant than in ‘ordinary’ information infrastructures (Klievink & Janssen, 2014).

The actors involved in the PII will attempt to influence decision-making to pursue their own goals and visions. We define actors as representatives of (parts of) an organisation involved in or a stakeholder of the data exchange process. We explicitly specify parts of an organisation, as different sections may be involved in the PII for various reasons, each with its own goals and values. Actors primarily collaborate to serve their own interests (Selsky & Parker, 2005). This is why public-private partnerships need to balance the values and objectives of both businesses and public organisations (Linder, 1999). These goals and visions do not always align with public values or the values of government organisations. The public domain is already characterised by conflicting values, such as dilemmas between efficiency and equity (Le Grand, 1990) and efficiency and democratic legitimacy (Weihe, 2008). In information exchange processes, many values must be protected by governments.

Governmental values may include transparency, legitimacy, a level playing field, and quality in public service delivery. The Dutch Government's data agenda (2020) states that in the Netherlands, the government is responsible for ensuring transparency in its participation, strengthening democracy, and promoting accountability to its citizens. Dutch law imposes strict requirements on information management, including the protection of personal data and the maintenance of system and process integrity. In recent years, there have been increasing demands for public, transparent, and traceable decision-making.

The complexity of sharing information increases as laws and regulations governing information collection, use, and sharing become more stringent. For example, the General Data Protection Regulation

(GDPR) or in Dutch, AVG, requires data processing to use only the data necessary to accomplish a specific task (data minimisation principle) (European Parliament & Council of the European Union, 2016). Moreover, information gathered for one purpose cannot be repurposed without explicit consent (purpose limitation). At the same time, the government must adopt new technologies to create more efficient systems. Some of these values conflict. The technological complexity grows as systems become increasingly opaque to many users, making accountability and transparency more challenging. It will also be harder to regulate external organisations that aim to uphold the government's values.

The social components within the PII become more complex to manage as processes, rules, agreements, and human resources are shared or distributed among different stakeholders. These components and the decision-making processes related to them are subject to negotiation and interaction. This will make it more difficult for governments to shape PII to suit their preferences.

3.1.4 SOCIO-TECHNICAL SYSTEMS

The fourth characteristic of an information infrastructure is that it is a socio-technical system. Socio-technical systems cannot be maintained or developed solely by focusing on either the social or the technical components. Human interaction critically influences the technical design (Ottens et al., 2006), and the technical design, in turn, shapes social interactions within the system. A socio-technical system must be researched and maintained as an integrated whole.

An information infrastructure is a combination of, on the one hand, technical components such as hardware, software, and networks, and, on the other hand, a social component involving stakeholders who provide these technologies and the relationships among them (Cordella, 2010; Braa et al., 2007). The only way to achieve information exchange is by combining these two components (Bostrom & Heinen, 1977).

The interdependence between single information systems is expanding, making it impossible to observe them as distinct from Information infrastructures (Hanseth, 2010). *“II are larger and more complex systems, involving large numbers of independent actors as developers as well as users.”* (Hanseth & Monteiro, 1998, p.49). The information infrastructure is not functional without stakeholder interaction in its operation and use. As a result, the dependency on another organisation increases. As Janssen et al. noticed: *“Many public organisations ignore the fact that they are dependent on others in their development projects and that needed functionalities might already have been developed by other organisations.”* (Janssen et al., 2009, p.234). Since multiple parties provide components of the information infrastructure, the dependency on other organisations increases. If one service provider fails, the whole infrastructure might fail (although this depends on the design).

What is unique about collaborations in PII is the potential for horizontal relationships. Research indicates that new digital government infrastructures often lack a central authority (Janssen et al., 2009). All organisations are interconnected, forming an information exchange chain that creates high interdependencies within a PII. Private organisations rely on governments for requesting and sharing information services through market creation, providing resources, and establishing laws and legislation. In turn, governments depend on private organisations to deliver services within the PII. Having private

organisations provide services can accelerate technological progress in components, reduce the risk of development failure, and require fewer resources. However, it may also reduce government control over the infrastructure. When the information infrastructure is decentralised and no longer controlled by information-requesting governments, those governments rely on other organisations to meet their needs. The dependence on external organisations for high-quality information and distributed infrastructure components is growing. This could lead to a 'black box' scenario, where governments utilise infrastructure components for information exchange without knowing the actual content of those components.

For public organisations, governing the use of PII is becoming increasingly complex. The black boxes may require stricter control over compliance with regulations and requirements. Not all decisions are made centrally within a single governmental organisation; therefore, governments have limited influence over the components within the information infrastructure. Hidden complexities can lead to failure, as components become interdependent. Governmental organisations need to determine their configurations, which consist of a mix of components. Creating and managing an infrastructure that delivers public value while motivating private parties to contribute and maintain continuity is a complex task for governmental organisations. This involves managing multiple systems and stakeholders. However, knowledge on how to achieve this remains scarce.

3.1.5 EVOLVING INFORMATION INFRASTRUCTURES

The fifth characteristic is that an information infrastructure is not static but constantly evolves in response to the decisions made by the actors involved. An information infrastructure has an emergent nature and is continuously developing (Cordella, 2010). *“Information infrastructures grow and develop over a long period of time, new parts are added to what exists, and existing parts are replaced by improved ones?”* (Hanseth & Monteiro, 1998, p. 49). Infrastructures are constantly evolving and require adjustments from those designing them. For example, when new components become available, this often results in more or new types of information, requiring more resources to process and manage (Bowker et al., 2010).

We established that, to some extent, the public information infrastructure is becoming more open. An open information infrastructure, according to Hanseth & Monteiro (1996), *“is built through extensions and improvements of what exists- never from scratch. It is open in the sense that any project, independent of how big it is, will just cover a part of an information infrastructure. The rest exists already and will be developed by others, being out of reach for the project and its control.”* (p.49). This is due to legacy, as *“Legacy systems tend to have long lifespans, as they often are too expensive or critical to replace”* (Matthiesen & Bjørn, 2015, p.876). Legacy is a major issue in government systems, as past choices influence future decision-making possibilities. The study by Matthiesen and Bjørn (2015) demonstrates that the legacy of government information infrastructure is often embedded within other frameworks, including laws, legislation, policies, and interconnected IT systems. Even when an organisation chooses to discontinue all systems, every subsequent decision leaves a legacy for the future.

However, in the context of a publicly governed and designed information infrastructure, the degree of openness of a PII depends on the choices made by the authorities involved, who are often policymakers or

receiving organisations. They can diverge from legacy systems and develop new information infrastructures. Yet there is a tendency, although it remains a choice, to reuse existing standards and parts of existing information infrastructure. Particularly on the information-providing side of data exchange, existing technology and systems are reused. Furthermore, there is a trend towards increased openness of information infrastructure in the public sector, where multiple authorities are involved in the design, maintenance, and adaptation of a PII. Together, these authorities develop the PII across the various stages.

We argue that the traditional literature on infrastructure lifecycle models can complement the installed base from the information infrastructure literature (Cordella, 2010). In classic infrastructure lifecycle models (Bolton & Foxon, 2015; Edwards et al., 2009), an infrastructure can be distinguished into multiple phases. This model encompasses the four lifecycle phases of *'emergence and establishment'*, *'expansion and momentum'*, *'stagnation and inertia'*, and *'transition and renewal'*, which in turn evolve into *'emergence and establishment'* as the renewed part of the infrastructure is implemented within the existing structure. The infrastructure will continue to develop throughout its lifecycle, provided that energy is invested in it. The lifecycle depicted in the literature is visualised in Figure 3-1.

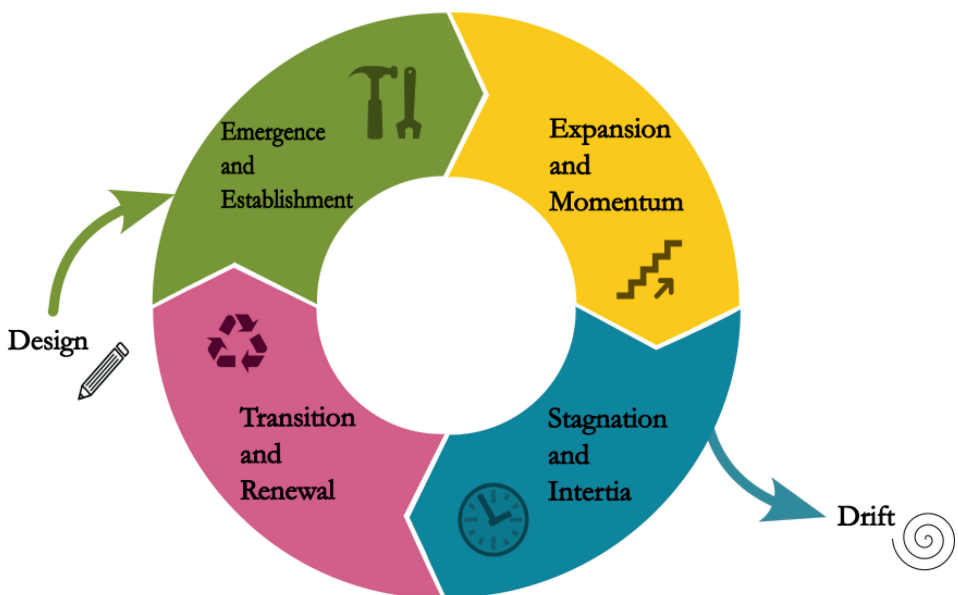


Figure 3-1. Lifecycle of an information infrastructure, visualised model of the theory by Bolton & Foxon, 2015; Edwards et al., 2009

First, the emergence and establishment phase. A new type of information infrastructure starts with a design. In this design, parts of the (public) information infrastructure can be designed from scratch, after which they evolve (Janssen et al., 2009). Over time, however, it is less likely that an entire information infrastructure will be shaped from scratch, as reuse of existing technology is common. Still, new services and functionalities of an information infrastructure are developed and implemented for information exchange. In this initial phase, new systems, technological components, or data-sharing processes are

introduced, and foundational infrastructure elements are implemented. According to Hanseth & Lyytinen (2010), information infrastructures are often developed through organisations experimenting and discovering new uses for existing techniques. There is a high level of uncertainty as actors define basic processes and technical components differently. Research by Bharosa (2022) suggests that small technical solutions must mature before they become mainstream. They need to demonstrate their value before they can be applied across multiple organisations. Promoting the use of open standards can enhance interoperability (Bharosa, 2022) and facilitate the II's progress towards the second phase.

Second, there is the phase of expansion and momentum. As adoption increases, standards begin to stabilise, and a prevailing configuration emerges. The infrastructure becomes more widely accepted among stakeholders in the information infrastructure. The value of a standard lies in its adoption. As the number of users of a standard grows, the value of the standard also rises. Additionally, the increased adoption of popular standards makes the information infrastructure more attractive to use (Hanseth & Lyytinen, 2010). Over time, technological components and data-sharing processes become more integrated into the existing information infrastructure and broader organisational processes. Towards the end of this phase, established stakeholders can exert their influence in the information infrastructure, which can hinder further development and innovation (Edwards et al., 2007).

Following this phase is the phase of stagnation and inertia. Over time, the infrastructure experiences extensive use; however, it begins to show signs of ageing. At this stage, the infrastructure is primarily embedded in existing processes and may therefore become resistant to further development. "*Organisational routines are highly efficient modes of organising behaviour, but they can be strongly resistant to change.*" (Edwards et al., 2007, p.17). According to Edwards et al. (2007), due to path dependency, some choices become irreversible. True path dependency exists when an alternative path would be more efficient in the new situation. It creates a legacy that limits further development of the information infrastructure. Concerns may arise regarding scalability, technical debt, or misalignment with current needs.

The final phase is 'Transition and Renewal', during which existing infrastructure either adapts or fails to do so. "*Long-standing infrastructures turn out never to be quite as finished as we might suppose.*" (Edwards et al., 2007, p. 24). In response to the limitations of phase three, infrastructures can evolve through upgrades, the integration of new standards or technical components, or governance reforms. Alternatively, the information infrastructure may be gradually replaced by a new design. The challenge during this phase is to determine who 'owns' the infrastructure to make this decision (Edwards et al., 2007). This ownership can result from a design choice or involve negotiation among various stakeholders. In each phase, stakeholders influence the design and reshape the installed base. Infrastructures are not linearly developed or static objects that can be implemented (Star & Ruhleder, 1996). Information infrastructures are developed through changes to the installed base. According to Tilson (2010), "*only a stable installed base allows new connections to be created*" (p.754). Often in an information infrastructure, a pre-existing environment exists before the data exchange is created (Aanestad et al., 2017). They observed that: "*To manage the further evolution of the installed base is challenging, as it entails building on the installed base and transforming it at the same time.*" (2017, p.30). Though redesign can be challenging, it is essential for the existence of the PII and preventing it from drifting.

During the design and redesign phases, the installed base is adapted. The introduction of new information infrastructures inevitably requires adaptations to existing information infrastructures and systems. For the adaptation, new standards can be developed to establish interoperability in the information infrastructure. As Cordella (2010) noted, “*Adaptations, ad hoc interventions, and bricolage are needed to keep the infrastructure alive, but they take time to be devised and implemented.*” (p.35) These efforts increase technological complexity, making subsequent changes and adaptations more challenging. Over time, the installed base becomes more deeply embedded, reinforcing current structures and guiding the infrastructure towards the third phase of its life cycle: stagnation and inertia. Beyond technical modifications, infrastructure changes also impact its organisational component. The introduction of new processes, systems, or standards often reshapes roles and responsibilities among stakeholders. As Cordella (2010) highlights, “*Every change in technology or in the organisation of work activities produces side effects.*” (p.34). These side effects may be intentional or unexpected. Depending on how they perceive the impact, stakeholders may embrace the changes if they are beneficial or oppose them, as they become ‘angry orphans’ (Hanseth & Monteiro, 1998). Over time, technologies and processes may start to diverge from the original design objectives of the information infrastructure, a phenomenon described by Ciborra (2004). This drift can ultimately lead to the replacement of the existing infrastructure by a new one. Therefore, infrastructure should be regarded as a dynamic system—constantly evolving and never fully stabilised.

Stakeholders play a crucial role in shaping an information infrastructure. At each stage, they can influence the design of the PII. We notice that design choices are formulated during the emerging and renewal phases, that the number of stakeholders grows during the expansion and momentum phases, and that, without initiation and resource allocation by the stakeholders, the information infrastructure may lose direction. Cordella (2010) describes circularity as an essential characteristic of an information infrastructure. “*Circularity is a relationship where every actor network affects and is affected by the characteristics of the actors, and where emergent characteristics of the actors affect and are affected by the characteristics of the actor network.*” (Cordella, 2010, p.35). There is a high level of interconnectedness, which can make some parts irreversible. Especially in governmental organisations, this means that choices have to be considered upfront. Once a standard or technical component is embedded in the information infrastructure, it can be costly or time-consuming to alter the standard.

It is therefore important to understand the design space and the underlying design choices. There is a need to maintain an overview of the ongoing development of the PII concerning data, technical, and governance quality. In this paper, we identify PII stewardship as a role that can fulfil this function. The following section further elaborates on the concept and definition of PII stewardship.

3.2 DEFINING PII STEWARDSHIP

As shown in previous sections, PIIs entail both social and technical components. Both components are subject to change: new technical standards and new laws are drivers for change, i.e., the evolution of PII

components. Technical components provide functionalities (e.g., for application integration or authentication). While functionalities are often stable or new functionalities are required by new regulations (e.g., attribute sharing with eIDAS wallets), the standards and technical specifications are subject to continuous innovation on a global level (Bharosa, 2022). Accordingly, governments need to source new technical components regularly. The sourcing of technical components can be achieved in multiple ways, including internal developments, tendering, public-private collaboration or other social structures such as trust frameworks (e.g. Van der Peet, 2025). Consequently, PII consists of interrelated technical and social components. As shown in Chapter 1.2, as the PII evolves, the complexity of PII increases. The challenges for governments involve managing increased reliance on external private organisations, navigating conflicting values among organisations, overseeing multi-stakeholder decision-making, and addressing rising technical complexity. Current literature offers limited guidance on how to steer PII and handle the numerous design choices that influence the PII quality. Here, we introduce a new concept of public information infrastructure stewardship, offering a promising perspective for exploring how PII can be designed, maintained, and adapted to enhance quality, long-term usability, and alignment with public values. Theoretical concepts such as ‘governance’ and ‘data stewardship’ alone are insufficient to fully grasp the authorities involved. However, they provide a starting point for understanding these new PII authorities.

In the following sections, we will discuss these terms and assess their relevance to PII stewardship.

3.2.1 GOVERNANCE

Governance in this research is applied in two ways: first, to understand how it relates to PII stewardship; and second, as a component of PII, as presented in Chapter 3.2, including governance structures. In Chapter 5.3, we focus on specific governance structures in a PII. However, in the following paragraphs, we examine the concept of governance to understand how this can contribute to understanding PII stewardship.

The concept of governance is frequently discussed in the literature on information infrastructures. It originally derived as a synonym for the word "government" (Stoker, 1998). Over time, the definition of governance has evolved to refer to governing styles in both the public and private sectors (Stoker, 1998). However, this concept has been applied so broadly and variably that there is no longer a consensus in the literature about what governance truly is. As a result, the concept no longer specifically pertains to the division of roles necessary to steer a PII towards high quality. Building on Stoker (1998), governance can be understood as the process of defining and distributing responsibilities for actions and decisions. Authorities can, however, be distributed, resulting in a chain of decisions; it is the governance that defines and allocates the decision-making authorities (Janssen, 2025). In the literature, a distinction is made between structural, procedural, and relational governance mechanisms (Peterson, 2004). Structural governance mechanisms comprise the structures and bodies that define roles and responsibilities. Relational governance mechanisms facilitate collaboration among the actors involved in governance, while procedural governance mechanisms encompass the rules, processes, and formal agreements (Peterson, 2004; Schneider et al., 2023). All these mechanisms can be utilised to establish the governance of, in our case, the PII. This research focuses on

governance of technology and the authorities involved with the formal decision-making rights. Building on existing literature, we identify four characteristics relevant to decision-making in PII, which can be used to develop a tentative definition of PII stewardship.

The first characteristic of governance is its collaborative nature, involving multiple actors (Bovaird & Löffler, 2003; Stoker, 1998; Klievink & Janssen, 2014). Information infrastructures can span organisational, sectoral, and even national boundaries. Second, public sector governance is often rooted in a formalised structure, laid down in agreements (Klievink & Janssen, 2014; Weill & Ross, 2005). These formal structures ensure stability and predictability, and enable authorities to make formal decisions. This means that the shared use and development of infrastructure follow clear principles rather than being decided on an ad hoc basis.

Thirdly, governance provides oversight of the information infrastructures and defines what must be done, who has the authority to make decisions, and how accountability is organised (Stoker, 1998; Khatri & Brown, 2010; Klievink & Janssen, 2014; Weill & Ross, 2005; Wende, 2007). As stated by Klievink and Janssen (2014), governance is “*defining and allocating of actions and decisions that ensure a form of collaboration that cannot be externally imposed*” (p.241). In other words, governance provides insight through a framework that delineates responsibilities, decision-making processes, and control mechanisms.

Finally, governance is typically organised around a standard, system, or process (Khatri & Brown, 2010; Thompson, 2003; Weill & Ross, 2005), and refers to the management and regulation. Governance is often established to create common ground among diverse stakeholders or initiated by specific institutions, such as governments or regulatory bodies, to address a specific problem or coordinate agreements around a particular system. In such cases, governance does not necessarily cover the entire information infrastructure but instead concentrates on a specific component or domain. An example here is the global governance of the XBRL standard used for the Dutch Standard Business Reporting PII (see section 4.1.1). Over time, however, these more localised arrangements can evolve and expand, gradually developing into a broader form of information infrastructure governance (e.g. a national taxonomy based on the XBRL standard).

These characteristics demonstrate that governance in information infrastructures is collaborative, yet formalised and structured. This simultaneously shapes the distribution of power and the coordination of practices within complex sociotechnical systems. However, the concept of governance alone does not provide sufficient insight into how PII are steered and the authorities involved in ensuring its quality. The authorities within a PII are not always concentrated in a single place; instead, they can arise across different layers and governance structures. This study focuses on identifying specific authorities and parts of organisations that contribute to maintaining and safeguarding quality within the PII. We draw upon the concept of data stewardship, which offers a well-documented perspective for understanding how responsibilities over the data and its process are important for the level of PII.

3.2.2 DATA STEWARDSHIP

The second concept related to the steering of PII is data stewardship. According to Rosenbaum (2020), data stewardship is: “*a collection of data management methods covering acquisition, storage, aggregation, and deidentification, and procedures for data release and use.*” Wendelborn (2023, p.3) defines data stewardship as the responsibility to optimise (research) data management to both enhance its potential value and, at the same time, protect values such as legitimacy and inclusion, and reduce asymmetry in the power balance. In the definitions of data stewardship, there is a focus on ownership, the development of data management and data sharing systems, and the quality management of the entire data process. This is closely related to the type of steering required for information sharing involving PII.

While governance entails the overarching rules and structures, data stewardship offers a more detailed perspective on data management responsibilities. In the definition of data stewardship, various roles are mentioned, ranging from ‘data custodian’ (Cheong & Chang, 2007) to ‘data quality manager’ (Khatri & Brown, 2010). The concept of data stewardship offers a potential solution to coordination challenges within information infrastructures by assigning responsibility for data within an organisation to a designated data steward role. As defined in research by Otto (2011), “*While data owners are ‘accountable’ for the immediate correctness and consistency of certain data, data stewards develop and provide the rules for the handling of this data, i.e., they are ‘responsible’ for the overall data management.*”(p.8). This means the person or organisation performing the tasks is responsible for protecting the information during the process and ensuring it is fit for use. In other words, data stewardship refers to a set of tasks focused on ensuring that the right data reaches the correct processes in the appropriate format and complies with applicable laws and regulations (Dawes, 1996; Dawes, 2010). More specifically, data stewardship explicitly links responsibilities to data (Rosenbaum, 2010). The role of the data steward is central to ensuring effective quality management (Wende, 2007). Data stewardship, therefore, indicates that a person or organisation is responsible for maintaining aspects such as accuracy, validity, security, integrity, usability, description, and long-term preservation of an organisation’s data (Dawes, 1996).

Data stewardship is crucial for responsible data management, ensuring trustworthiness, and supporting stakeholder objectives. A key task is documenting metadata or data definitions. Data and metadata standards, for instance, set how data elements are described, defined, and represented within systems. These standards improve data quality and facilitate the use of data for various purposes (Dawes, 2010). To ensure data quality, including accuracy and completeness, proper documentation is necessary (Wendelborn et al., 2023). This aids data integration and interoperability, particularly in environments with multiple stakeholders.

The responsibility extends not only to the quality of the data itself but also to its technical quality. While data quality or information quality refers to the characteristics of the system’s output, system quality (or, in our case, technical quality) refers to the quality of the information processing system itself (Delone & McLean, 1992). To avoid confusion, we refer to technical quality instead of system quality. Additionally, because in a PII, there are often multiple systems involved rather than a single system. Enhancing

information systems is often linked to improving technical quality (Gorla et al., 2010). A fundamental task of data stewardship is coordination between systems. This involves tasks such as establishing protocols to prevent the development of large static databases, which are vulnerable to tampering (Rosenbaum, 2010). It includes building trust and competence among stakeholders, creating privacy policies and access policies, supporting standardisation initiatives, and enabling seamless data exchange (Rosenbaum, 2010). Improved coordination helps establish trust for the individuals and entities using the system (Wendelborn et al., 2023).

Data stewardship is not just a technical or operational task, but a role that is carried out through effective governance and collaboration. At its core, data stewardship is a systematic, explicit responsibility to all relevant stakeholders (Wendelborn et al., 2023). This means a large part of the responsibility involves identifying stakeholders, ensuring that the objective is clear, and aligning priorities among different stakeholders. There should be transparency for all stakeholders to avoid ambiguity around data-related tasks. Additionally, data stewards facilitate lifecycle management by guiding the process from the design phase through implementation, expansion, and transition. Throughout these stages, the information must consistently remain of high quality.

Data stewardship plays a critical role in both public and private organisations, as they rely on data for decision-making, service delivery, and regulatory compliance. Stewardship entails the capacity to obtain, use, and administer information as a resource in accordance with organisational, jurisdictional, or societal value across purposes and over time (Dawes, 1996). *“A good quality level is fundamental in order to provide services in both public and private contexts.”* (Scannapieco & Catarci, 2002, p.1). Still, in the public sector, the stakes are even higher. Governments rely on accurate and timely data to formulate policies, allocate resources, and deliver services to citizens. Poor data quality or poorly managed information can result in inefficient public services, misguided policy decisions, and a decline in public trust. In 1996, a study by Sharon Dawes proposed the policy principle of data stewardship for government agencies seeking to increase openness and transparency by enhancing information access and dissemination (Dawes, 1996).

3.2.3 DEFINING PII STEWARDSHIP

It is observed that the concept of governance establishes higher-level authorities and their formal decision-making rights. In contrast, the term 'stewardship' refers to executive-level responsibilities for establishing quality in PII at an operational level. Moreover, as Dawes (1996) suggested, data stewardship requires the appropriate allocation of responsibilities. In practice, information infrastructures are becoming increasingly larger and more complex (Hanseth & Monteiro, 1998). As a result, the boundaries of an infrastructure can become blurred. Consequently, multiple governance structures may coexist within a PII. Furthermore, several researchers argue that information infrastructures cannot be fully controlled due to ongoing development and increasing complexity (Ciborra, 2004; Hanseth & Lyytinen, 2010).

This situation highlights the need for an authority capable of designing, maintaining and adapting the PII landscape. In complex PII, it can entail working across multiple layers and formal governance

structures, and aligning with multiple stakeholders within the PII. It ranges from operational execution to high-level insight into PII.

3.2.4 THE CONCEPT OF PII STEWARDSHIP

Building on the previous discussion, we introduce the new concept of Public Information Infrastructure Stewardship. PII stewardship extends the principles of data stewardship to encompass an entire infrastructure, defining the collection of authorities for maintaining the quality of PII. Building on data stewardship, these responsibilities include ensuring data quality, maintaining technical quality, and collaborating with relevant stakeholders or ensuring governance quality. Additionally, PII stewardship encompasses management during the entire lifecycle of an information infrastructure, from design and implementation to expansion and transition phases, ensuring quality is upheld throughout.

In addition to these governance- and data-stewardship-based characteristics, PII stewardship introduces three key extensions.

First, within the PII domain, the governance quality becomes increasingly important. Since a PII may be composed of multiple technical components, the number of involved actors usually increases, and various governance structures may coexist, responsible for parts of the PII. As a result, alongside data and technical quality, governance quality becomes a vital concern. Effective governance within a PII requires, for instance, transparent decision-making processes and cost-efficient management, as poor governance can significantly hinder the PII's functionality and development. PII Stewardship in this context, therefore, not only protects data and technical quality but also strengthens robust, well-functioning governance structures.

Secondly, PII stewardship authorities can be part of a single organisation or a collaborative of multiple organisations. While traditional data stewardship tends to be located within a single organisation, PII stewardship must accommodate the distributed and interconnected nature of public information infrastructures. Authorities installed for maintaining quality can therefore be shared among multiple organisations in different ways. This means that stewardship is no longer restricted to a single, centralised role but can take various forms, from a centralised approach similar to conventional data stewardship to shared or distributed PII stewardship. As with data stewardship, these roles may also be performed by third parties.

Finally, the form and focus of PII stewardship are inherently dynamic, evolving as infrastructures change, new technologies emerge, and governance needs shift. As theory shows, a PII is a constantly changing entity. Just as the configuration of a PII captures a moment in time, the stewardship of a PII is also subject to change. PII Stewardship can develop and adapt to reflect the evolving nature of the infrastructure.

Taken together, the concept of PII stewardship bridges the gap between operational data stewardship and high-level governance. The PII stewards are authorities who ensure that complex infrastructures are managed to a high standard throughout their lifecycle. Crucially, PII stewardship ensures the design and shape of the PII configuration itself, recognising that decisions made influence the quality of

the entire PII. As earlier shown in Figure 1-1, the PII configuration also shapes the possible PII stewardship. This means that decisions made by PII stewards can shape the PII configuration and, therefore, alter the PII stewardship.

Based on these considerations, we have a tentative definition for PII stewardship. *PII stewardship refers to the authorities for ensuring data, technical and governance quality in designing, maintaining, and adapting a public information infrastructure throughout its life cycles.* This stewardship must be exercised throughout the entire lifecycle of the PII, encompassing the design, implementation, expansion, and transition phases.

Based on the literature, we already observe that there are multiple types of stewardship in which one or more actors fulfil roles to ensure the quality of the PII. Due to the involvement of several actors in the PII, collaboration can be stated between different stakeholders for stewardship roles, whereas other PIIs might be managed by a single actor.

In the context of single-actor PII stewardship, the involved organisation is expected to play a central role in ensuring quality by defining a format, establishing technical standards, systems, and service delivery. Moreover, it is anticipated that they will establish a governance structure to support (bilateral) coordination among actors involved in PII management. On the other hand, in shared PII stewardship involving multiple actors, we expect multiple organisations to contribute to technical quality by collectively establishing standards, systems, and services, or by ensuring data quality through the creation of a data format. We therefore expect there to be different categories of stewardship, ranging from single-actor to collaboration-oriented stewardship. Chapter 7 elaborates on the empirically derived classification of PII stewardship.

3.3 QUALITY FRAMEWORK FOR PII

To assess the quality of a PII, what PII stewards strive for, it is first necessary to understand what constitutes PII quality. Based on the characteristics of a PII, three levels can be identified in its design, as presented earlier in Chapter 2.3.1 and Figure 2-2. The first level involves the data exchanged through the PII; this is the key element for which the PII is designed and is often formatted to enhance quality. The second level relates to the technical component, including the systems used for data exchange, the services offered, and standards such as security or exchange protocols. Finally, PII are socio-technical systems in which actors play a vital role; therefore, one or more governance structures are part of the overall governance configuration of PII. Considering these levels, it follows that the data, technical, and governance quality should all be high. There is no comprehensive list of measures available for assessing these quality aspects at the level of an entire PII. In their work on data quality and systems, Delone & McLean (1992) offer an extensive list of quality measures. Both data quality and system quality encompass multiple measures (for example, data quality might include over 30 measures), and it is not clear upfront which measures should be prioritised or included in a comparison framework. Conversely, there is currently no corresponding model for governance quality.

Therefore, the development of the PII quality framework also drew on insights from exploratory case study interviews. Through interviews with nine respondents, we examined the dimensions of quality in

PII. This study focuses on three empirical cases, as described in Chapter 4.1. The numerous measures identified for each category were used to design a semi-structured interview protocol. The resulting framework, which will be discussed in the following sections, emphasises the measures most relevant to data, technical, and governance quality in PII.

3.3.1 DATA QUALITY

Data quality refers to the quality of the information systems' output (Delone & McLean, 1992). There is no consensus on measures for the data quality dimension. Based on the identified measures in the literature and the interviews, we concluded that the measures of PII data quality are accuracy, completeness, consistency, relevance, timeliness, and format.

First, **accuracy** is defined as the extent to which the recorded value corresponds to the actual value in the real world (Fisher & Kingma, 2001). The dataset should be of high integrity, with no errors or errors that are easily identifiable (Wang & Strong, 1996). Especially in a PII, this can be challenging due to the many stakeholders involved in the process. Second, **completeness** refers to “*the degree to which values are present in a data collection*” (Fisher & Kingma, 2001, p. 110). An incomplete dataset has missing values. As more variables are missing, its overall completeness and quality decline. In a PII, when multiple stakeholders are involved, this issue can often be identified during the process if the format is properly filled out. Next, **consistency** can be linked to the definition of semantic rules for the data exchange (Batini et al., 2009), implying that the meaning of the data element aligns for all parties (Fisher & Kingma, 2001). Having more stakeholders in a PII can jeopardise consistency due to varied interpretations. Fourth, data **relevance** refers to the extent to which the user can actually use the data (Fisher & Kingma, 2001). This is often determined between the sender and receiver of the data. Next, **timeliness** is a concept that is the subject of much debate. In this paper, timeliness implies that the dataset is not outdated (Fisher & Kingma, 2001). In a timely dataset, real-world modifications are (directly) represented. This refers to the agreements made between the sending and receiving organisations. And finally, **fitness for use** refers to the **format** in which data is presented, and whether it suits the purpose of the data user (Fisher & Kingma, 2001). We consider format and fitness for use as related concepts. “*Format is related to the presentation layout of information outputs.*” (Gorla et al., 2010, p. 213). Formats are often aligned within the PII, but multiple formats can coexist either in parts or throughout the entire PII.

Allowing more systems in the PII environment can compromise the overall quality of the data. Any transformation of the data or other services that intervene in it may compromise data integrity. The more actors that engage with the data, the more effort is needed to maintain its quality and accuracy.

3.3.2 TECHNICAL PII QUALITY

The concept of technical quality in a PII remains relatively underexplored. As a starting point, we draw on the concept of system quality from the information systems literature. This will be used to further develop a technical PII quality concept. System quality can be defined as the performance characteristics of the

information processing system itself (Delone & McLean, 1992). Delone and McLean (1992) gathered several measures from literature that contribute to system quality, such as system flexibility, response time, accessibility, ease of use, reliability, sophistication, and system integration. However, while these concepts provide a useful foundation, the context of information infrastructures introduces additional complexity. An infrastructure typically consists of multiple systems delivering diverse services and often incorporates standards to enhance interoperability. Insights from our exploratory case study interviews further informed this conceptualisation. Respondents emphasised the importance of auditability and robustness for technical quality in infrastructures comprising multiple interdependent systems. This resulted in 9 measures of technical PII quality.

The first two new measures are auditability and robustness. **Auditability** is the ability to track whether someone or something has accessed and/or modified data (Khatri & Brown, 2010). Auditability is an indicator of the integrity of the information system as a whole. Auditability often means having accessible audit trails and log files for data exchanges and transactions. Especially in a chain of systems, audit trails enable the identification of weak points. **Robustness** was added because it was deemed that a system not only needs to be flexible to change but also be able to cope with external pressures and be resilient to environmental changes. Fricke & Schulz (2005) defined robustness as a “*system’s ability to be insensitive towards changing environments*”. Low robustness means a PII is not resilient and will fail under external pressure.

Next, the respondents mentioned the **ease of use** of the PII and its systems. This measure refers to the perceived user-friendliness of graphical user interfaces for end users who enter and share data. The PII often involves various user groups across data exchanges, intermediaries, third parties, or software providers, making it easier for end-users to use, but also enabling intermediaries to create easily usable technology for other intermediaries.

The fourth important measure is the flexibility of the PII and its systems. **Flexibility** refers to the ability to handle variations in the requirements of a business process (Gebauer & Schober, 2006). This includes the extensibility. Not having enough flexibility in the PII can limit its effectiveness, preventing its use in certain situations. In the end, this can result in a decrease in the lifetime (Furukawa & Minami, 2013). Fifth is the **response time** of the systems in the PII. This refers to “*the degree to which a system offers quick (or timely) responses to requests for information or action*” (Nelson et al., 2005, p. 201). The PII’s response time depends on the systems involved.

Next is the PII’s reliability. **Reliability** can be defined as two concepts: **availability** and **integrity** of the PII and its systems. Availability refers to the uptime (Nelson et al., 2005). This, too, depends on the availability of the systems in the PII. In contrast, integrity can be defined as trusting that the data exchanged through the PII is free of error and not manipulated (Närman et al., 2008).

The sixth measure is the sophistication of the PII and the systems. **Sophistication** refers to whether the PII and its systems use state-of-the-art technology. Sophisticated PII uses modern technology and features user-friendly interfaces (Gorla et al., 2010). The seventh measure is the accessibility to the PII. Accessibility refers to whether data is available when the user needs it. Nelson et al. (Nelson et al., 2005) define accessibility as “*The degree to which a system and the information it contains can be accessed with relatively low*

effort” (p. 205). The final measure of technical quality is **interoperability**. In PII, having interoperability is especially important. Interoperability is a multidimensional concept and can encompass technical, procedural, semantic, organisational, and legal interoperability. Systems created by different organisations should be able to interact with one another without limitations. Each limitation is a barrier to creating public value through a data ecosystem and prevents organisations from joining it.

We note that the technical quality of a PII is the sum of the quality of its constituent systems. Yet, depending on the configuration and the integration of the systems, a system can have more or less influence on the overall quality. Additionally, some qualities, such as flexibility and robustness, or traceability and ease of use, can lead to contradictions.

3.3.3 GOVERNANCE QUALITY

The third quality dimension PII stewards should strive for is governance quality. PII stewardship refers to the authorities for steering the PII, and they therefore need to ensure a governance sub-configuration for data exchange processes. “*Data governance refers to who holds the decision rights and is held accountable for an organisation’s decision-making about its data assets.*” (Khatri & Brown, 2010, p. 149). Where traditional single organisations have less fragmented responsibilities and accountabilities and easier control over data (Khatri & Brown, 2010), government-business collaborations experience greater delegation of authority, which can lead to unintended outcomes (Brinkerhoff & Brinkerhoff, 2011). Based on the literature and interviews, ten governance quality measures were identified.

The first measure is accountability. **Accountability** refers to actors being called to account to some authority for their actions (Bovens, 2007). This means organisations in the PII have to justify their decisions and actions. The question arises: to whom do they owe an explanation of their behaviour, and who can impose consequences? Governance includes defining, assigning, and implementing responsibilities (Rosenbaum, 2010). This relates to decision-making rights, which determine the extent to which actors may influence the final decision (Bruijn, Bruin, & Heuvelhof, 2008).

Secondly, governance needs to be transparent and accountable (AlShamsi et al., 2021; Devaney, 2016; Lockwood, 2009). In the governance context, **transparency** refers to greater visibility into decision-making processes for stakeholders. Additionally, it requires documentation of the authority's achievements and failures in governance (Lockwood, 2009). Next is the **adaptiveness and resilience** of the governance configuration. Resilience is the capacity of an entity to absorb change until it establishes new processes and structures (Lockwood, 2009). In a PII, governance configuration and its structures can be either resilient or adaptive to change, depending on its needs.

Next, the inclusiveness and openness of the PII. **Inclusiveness and openness** can be defined as enabling all stakeholders to participate in the decision-making process (Lockwood, 2009). Inclusiveness entails fairness, as those affected by a system should have a voice in its governance. Systems without inclusive participation are prone to transformation resistance and blocking power from those excluded from governance.

The fourth and fifth measures are **effectiveness** and **efficiency**. These are concepts related to output and outcome. “Both effectiveness and efficiency are valued intrinsically in public governance; they are values that constitute the core of public governance’s legitimacy” (de Graaf & Paanakker, 2015, p. 2). Efficiency is about achieving results (e.g., supported decisions, changes) with minimal resource expenditure (e.g., time and money) (Ernst, 2019).

Lastly, there are measures of clear funding and adequate **conflict resolution**. Several authors acknowledge the importance of resource allocation and funding (Organisation for Economic Co-operation and Development, 2002; Provan & Kenis, 2008), as well as clear conflict resolution (Provan & Kenis, 2008). Lacking clarity on funding or conflict-resolution procedures can lead to unmanageable governance structures. In particular, inter-organisational governance must include structures to resolve conflicts (Provan & Kenis, 2008).

Some of these measures are often specific to a single governance structure. However, a PII can also have multiple governance structures, and the quality of any single governance structure can influence the overall governance quality of the PII. The impact on overall governance quality varies by the size and significance of each governance structure.

Figure 3-2 presents an overview of PII quality, the three dimensions and their measures.

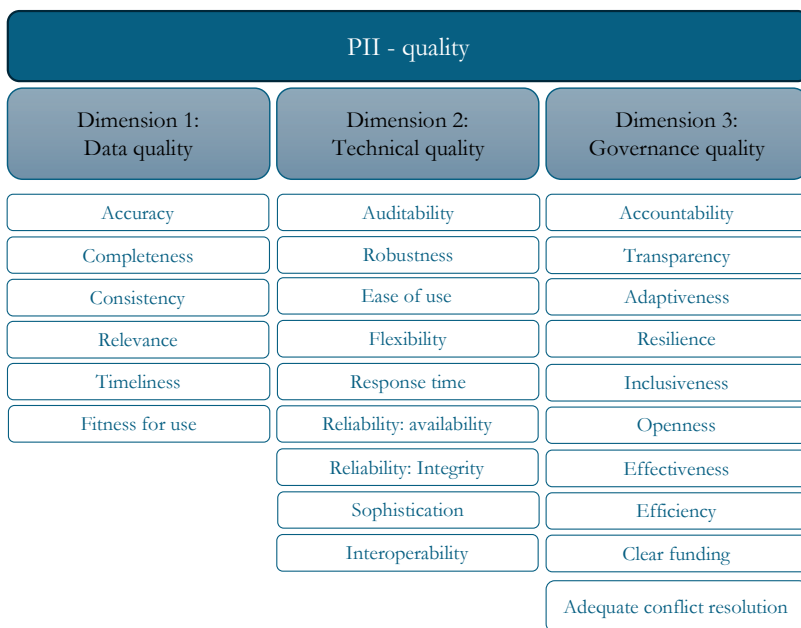


Figure 3-2. Overview of the PII quality dimensions and measures

This figure shows the preliminary framework for PII. A configuration of a PII can be evaluated on the PII quality measures, and it is the task of a PII steward to strive for these quality dimensions. In Chapter 6.7,

we assess the quality of the prevailing PII configurations, zooming in on some of the measures in this framework.

3.4 CHAPTER CONCLUSION

To conclude this chapter, we observe that public sector organisations require information beyond their internal operations and business processes, and need public information infrastructures to access it. Designing, implementing, and maintaining these structures and PII involves considerable complexity, as the decisions and interactions of the involved actors shape these infrastructures. For instance, due to technical development, collaboration, the adoption of standards, legacy issues, and an increase in stakeholders, complexity increases.

In this chapter, we defined public information infrastructures and PII stewardship. A PII is an information infrastructure explicitly designed for sharing information with the public sector. It is a socio-technical system that is not static but constantly evolves based on the decisions of the stakeholders involved in the PII.

To address these challenges and complexities, effective steering is essential. Building on the concepts of governance and data stewardship, we introduce a new form of steering for PII: Public Information Infrastructure Stewardship. We define PII stewardship as follows: *PII stewardship refers to the authorities for ensuring data, technical and governance quality in designing, maintaining, and adapting a public information infrastructure throughout its life cycles.* Until now, there has been no clear definition of the collection authorities guiding and ensuring the quality of PII concerning data, technical, and governance quality. PII stewardship offers a new definition that connects operational data stewardship with higher-level governance. We also present a preliminary framework for PII quality that PII stewards aim to achieve.

In the following Chapters, we provide deeper insight into the cases of PII and the contextual factors PII stewards face when shaping a PII.

Chapter 4

CASES OF PUBLIC
INFORMATION
INFRASTRUCTURES

This chapter discusses the cases that are central to the research. The case study approach provides an important foundation for understanding PII configurations and stewardship. Three cases have been examined in-depth and are exploratory in nature for both the phase one *Definitions* for establishing the quality measures as presented in Chapter 3.3 and for identifying PII design choices in phase two *Design variables*. These cases provide depth and a rich perspective on the diversity in design choices and quality measures.

In phase three: *Configurations*, elaborate on the different PII configurations. A total of twenty-two cases have been analysed for this purpose. This broader set was used to identify patterns and to determine prevailing configurations. In this chapter, we provide an overview of these cases. A summary of the case is available in Appendix B.

4.1 EXPLORATIVE CASES

To identify and analyse practical design choices in PII and to gain insight into PII quality measures, this study draws on three distinct case studies. These cases reflect different aspects and provide valuable insights into how PII are shaped in practice. Based on the selection criteria outlined in Chapter 2, we selected three cases of data exchange with governmental organisations. The data demanded by government agencies is essential for carrying out their primary functions and fulfilling their public duties. All cases were sufficiently similar to identify meaningful patterns, yet diverse enough in content and context to enable a constructive comparison. The selected cases include tax declarations via Standard Business Reporting (SBR), Customs declarations via the Single Window (SW), and eInvoicing to the central government. The selected cases involve data exchange between organisations and government agencies (X2G). The systems are data-driven and essential for the execution of public tasks, involving several actor groups such as government agencies receiving data, businesses providing data, and intermediaries.

In the following sections, we describe each case to establish the foundation for the analysis of quality measures in Chapter 3.3 and for the design variables described in Chapter Five.

4.1.1 TAX DECLARATION THROUGH STANDARD BUSINESS REPORTING

The first case used is the Tax Declaration through the Standard Business Reporting (SBR) system. SBR is the Dutch national standard for business-to-government financial reporting (Bharosa, Hietbrink, Mosterd, & Van Oosterhout, 2018). The government introduced SBR to simplify business-reporting obligations. In 2019, approximately 44,4 million messages were sent through SBR. SBR has three foundational components: the Dutch Taxonomy, the use of XBRL, and the shared digital gateway infrastructure for public agencies.

As a PII, tax declaration through SBR has a solid legal foundation. Information providers are required to submit data to the Tax Authority. The Tax Authority is responsible for offering suitable (electronic) channels (Wet moderniseren elektronisch bestuurlijk verkeer, 2023). Reporting data through SBR is unambiguous and reusable for various types of business reports (e.g., tax, annual reports, statistics reports) by standardising the labels and language used in business administration.

Data quality and system quality are crucial because there are legal consequences if data is not supplied to government agencies within the required timeframes. Therefore, SBR employs many data validation methods, including non-reputable electronic signatures, data validators, timestamps, and an audit trail system that retains data for up to seven years. Multiple organisations collaborate to provide the necessary functions for preparing, identifying, validating, exchanging, and processing information. Software developers embed SBR rules into their business and accounting software to make them ‘SBR-enabled.’ Businesses using SBR-enabled software can then report using information that has already been recorded during their operations.

SBR utilises international open standards, including XBRL and web services, to facilitate a high level of automation in the business reporting process.

Table 7. Facts and numbers of SBR to Dutch government organisations based on internal documentation. Taken from Van Donge (2022), p.4.

Type of data	Business reports
Number of messages	51,5 million (2021)
Message types	53
Stakeholders	<ul style="list-style-type: none"> • Governmental organisations as receivers: Tax authority (BD), Chamber of Commerce (KvK), executive education service (DUO) of the Ministry of Education, Culture and Science, The Housing Corporation Authority, The Social Housing Guarantee Fund, and The Ministry of the Interior and Kingdom Relations (together in the project ‘SBR-wonen’) • Data suppliers: e.g., Banks, taxpayers, schooling systems, housing corporations. • Intermediaries • Software developers • Accountancy organisations • Governmental Shared service provider: Logius • Policy: The Ministry of the Interior and Kingdom Relations, the Ministry of Finance, and the Ministry of Economic Affairs. • Other organisations: XBRL-Netherlands, NLdigital, Council for Annual Reporting.
Gateway to government	Digipoort, run by government organisations.
Data Governance	Participated by stakeholders and stakeholders’ representatives

The Dutch government developed a framework of standards and agreements for data exchange, in collaboration with accountants, software providers, and banks. Alongside representatives from various stakeholder groups (i.e., audit firms, bookkeepers, software vendors, and banks), the Dutch Government acts as the SBR governance body. It offers resources for further developing the information infrastructure (Bharosa et al., 2018).

4.1.2 CUSTOMS DECLARATION THROUGH SINGLE WINDOW

The second case used for research is the clearance process of Customs declarations via the Single Window (SW). This PII connects businesses and government organisations involved in international trade. Importing and exporting companies utilise the Port Community Systems (PCS) to send electronic declarations to Customs. These PCS often combine both public and private information. This declaration is obligatory for

trading organisations, as Customs needs to know what a vessel contains, where it has been loaded, and who is onboard. These declarations form part of the procedure to confirm that the cargo can enter or leave the port.

To comply with European Directive 2010/65, several European countries participated in the ‘Advanced National Network for Administrations’ (AnNa) project from 2012 to 2015. Here, the foundations were laid for the Maritime Single Window, now known as the Single Window, as air cargo messages are also included. The Single Window information infrastructure is developed to exchange maritime and air notifications with Customs and the Royal Netherlands Military Police. The aim of the Customs declaration (CD) through the Single Window is to establish a single point of contact. This means data providers report data only once. The SW will duplicate messages for multiple information requestors. One message enters the SW, which converts it into twenty-two message types, each for one or more recipients. This data includes details such as the arrival of ships or aircraft, cargo, onboard personnel, and Customs declarations.

Table 8. Facts and numbers of Customs Declarations to Dutch government organisations based on internal documentation. Taken from Van Donge (2022), p.4.

<i>Type of data</i>	<i>Information on, e.g. the arrival of ships or aircraft, the cargo, the people on board, and Customs declarations.</i>
<i>Number of messages</i>	47 million (2018)
<i>Message types</i>	22
<i>Stakeholders</i>	<ul style="list-style-type: none"> • The suppliers of data: captains and commanders, shipping companies, shipping agents, airlines, air cargo handlers, ship suppliers, provisioners, and bunkerers. • Port community systems, intermediary, hubs; providing IT-solutions for communicating with the government. • Data receiving Government organisation, Customs Tax authorities (Belastingdienst Douane), Rijkswaterstaat, Royal Netherlands Mare Chaussee. • Service providing government organisations, Customs Tax authorities (Belastingdienst Douane). • Policy, Ministry of Infrastructure & Water Management • Port authorities • European Committee
<i>Gateway to government</i>	Single Window, run by government organisations
<i>Data governance</i>	Participated by receiving parties

Data suppliers must provide information through an intermediary such as a Port Community System, a regional hub, or a web portal connected to SW. This data is often time-sensitive. If the data is not supplied, an aircraft or vessel cannot enter. Therefore, system availability must be high. Several stakeholders are involved in this process, as shown in Table 8. Not all stakeholders are involved in governance. Public governance is managed by the Ministry of Infrastructure and Water Management (I&W) and includes the receiving parties. This governance is primarily established to align requirements and ensure the quality of data and systems.

4.1.3 EINVOICING TO THE CENTRAL GOVERNMENT

The third case involves eInvoicing messages from businesses to governmental organisations. In eInvoicing, there are three methods for supplying data. The first is through a direct link with Digipoort. Large organisations maintain a direct connection with the central government message system Digipoort. Some of these organisations, with a direct link, have assumed the role of intermediary for other organisations. Software providers offering accounting systems facilitate connections with Digipoort to support procurement messages for Dutch government agencies.

Another way to connect with the Digipoort is through portals. Until the beginning of 2023, DigiInkoop, provided by the public intermediary Logius, offered a human-to-system portal for supply procurement messages. Since 2023, this has been decentralised into multiple different systems.

Table 9. Facts and numbers of eInvoicing messages towards the Dutch government in 2022, based on Internal documentation. Taken from Van Donge (2022), p.4.

Type of data	eInvoicing messages
Number of messages	1.7 million (2022)
Message types	10
Stakeholders	<ul style="list-style-type: none"> • Policy: The Ministry of the Interior and Kingdom Relations. • Governmental Shared service provider: Logius • Supporting organisation: Netherlands Enterprise Agency (RVO) • Independent supervision in the Netherlands: Radiocommunications Agency Netherlands (Agency of the Ministry of Economic Affairs and Climate Policy) • End-users (government and Businesses) • Software developers • Access points • OpenPeppol foundation • Peppol Authorities (Netherlands: NPa fulfilled by Ministry of the Interior and Kingdom Relations, Logius, and Netherlands Enterprise Agency) • European committee (CEN)
Gateway	Decentralised through the four corner model, interaction provided by access point/service providers, and Centralised through Digipoort, managed by government organisations.
Data governance	Participated by stakeholders and stakeholders' representatives

Finally, the third method of delivering eInvoicing messages is through the Peppol Network. Peppol utilises the four-corner model, which involves two service providers (corners 2 and 3) managing the information exchange on behalf of the information provider and receiver (corners 1 and 4). The Peppol network is part of the Peppol Association, which establishes agreements, document standards, and a legal framework that governs the network. OpenPeppol and the Peppol Authorities are responsible for creating and overseeing this legal framework. Since October 2020, the Dutch government has become a Peppol authority and established the Netherlands Peppol Authority (NPa). Unlike Standard Business Reporting and Single Window, Peppol lacks a legal foundation in the Netherlands.

There are several governance structures for eInvoicing. Internationally, CEN working groups are responsible for establishing European Norms for eInvoicing messages. Nationally, a NEN working group exists, as well as governmental regulations.

4.2 DESCRIPTIVE CASES

This section introduces the twenty-two cases used in the analysis to identify the configurations. These cases constitute the empirical basis for the study, and the results of the analysis are presented in Chapter 6.

For this research, the Customs declaration and eInvoicing cases for government organisations were reused. The initial aim was to include the case study of the Tax Administration. However, due to limited availability to participate in the research process, this case was ultimately not pursued further. Instead of the Tax Administration case, the participants in the SBR-education and WOCO cases, both active in the SBR trust framework, agreed to contribute to this study.

To improve clarity and readability, abbreviated or widely recognised names are used to refer to the cases throughout this chapter. It is important to note, however, that these names represent only a specific aspect of the broader information infrastructure. For example, the case referred to as ‘OOTS’ involves the retrieval of diploma data. While OOTS itself facilitates various types of data exchange, we use the term here to denote the diploma-related implementation, owing to its conciseness and recognisability.

The following Table 10 provides an overview of each case. Appendix B offers further insight for each case.

Table 10. Case description of the cases used in phase 3

Case	Case name used in research	Case description (type of data, sender, and receiver)	Stewardship: collaboration
1	Pension funds	Data from UWV to pension funds to report which citizens have fallen ill, which may affect their pensions.	Single actor
2	BRP issue reporting	User reporting issues to the registry of persons at the National Office for Identity Data (RvIG), which manages the registry.	Fragmented, but with an extensive central collaboration
3	eInvoicing	Invoice message from businesses of governmental organisations to central public organisations	Fragmented collaboration
4	Bank supervision	Banks in the Netherlands submit their financial reports to the central bank, the Dutch National Bank (DNB).	Single actor
5	SBR-education	Business reports from educational organisations to OCW and the Ministry of Education.	Fragmented collaboration
6	eCodex	Cross-border data exchange on legal documents with organisations such as CJIB (Centraal Justitieel Incassobureau), courts, public prosecutors' offices (OM), and DJI.	Central collaboration
7	Edu-V	Procurement of digital learning materials in Dutch primary, secondary, special, and vocational education	Central collaboration
8	Customs	Import declaration of goods (for the purpose of calculating import duties, fees, and checking of goods) from importing organisations to the Dutch Customs authority.	Single actor
9	WMO	Digital communication between municipalities and care providers under the Social Support Act (Wmo).	Fragmented collaboration
10	Child authority	Data on registered individuals, court decisions, and police records, which are available upon request from the authority module for the State Police.	Fragmented, but with an extensive central collaboration
11	Vaccination	Digital data exchange between the Youth Health Care (JGZ) and the National Institute for Public Health and the Environment (RIVM)	Fragmented collaboration
12	Land registry	Providing a registration of the purchase agreement and deed of transfer from a notary to the Land Registry.	Single actor
13	Kik-V	Business administration of nursing homes reporting to healthcare providers, regulators, and policymakers via the KIK-V infrastructure.	Central collaboration
14	National housing progress	Providing data on the progress of national housing construction from municipalities to the Ministry of Public Housing and Spatial Planning.	Fragmented, but with an extensive central collaboration
15	Wage returns statistics	Wage declaration information provided by the Tax and Customs Administration to the unemployment office (UWV) and the central agency for statistics (CBS).	Single actor
16	Migration reporting	Migration reporting from involved organisations (such as IND) to the Ministry of Justice and Security	Single actor
17	Carecodex	Carecodex, or BabyConnect, specialises in enabling digital data exchange between various healthcare providers, such as midwives, gynaecologists, maternity services, youth health care, and paediatricians.	Central collaboration
18	OOTS	Exchange and verification of diploma information between national authorities (for example, DUO) via the OOTS infrastructure.	Central collaboration
19	Single Window	Data exchange between businesses and government agencies during importing, exporting, and transiting goods. Reporting includes transport and logistics data, licensing, certification, and information on a ship's crew.	Single actor
20	VUM	Data exchange between public and private parties in the work and income sector aims to match job seekers to suitable vacancies more quickly and effectively by enabling standardised, secure data exchange.	Central collaboration
21	WOCO	Data exchange between municipalities and the Unemployment Insurance Agency (UWV) within the work and income sector. The aim is to match job seekers with suitable vacancies more quickly and effectively by enabling standardised and secure data sharing.	Fragmented collaboration
22	Notification property value	Municipalities and other government bodies exchange digital data updates on the assessed value of properties.	Fragmented, but with a large central collaboration

4.3 CROSS-CASE CONTEXTUAL FACTORS

Across the three explorative and twenty-two descriptive cases, we identify contextual factors that influence the design and stewardship of PII. These contextual factors were identified by systematically comparing multiple cases during the second and third research phase, as described in chapters 2.4 and 2.5, using a coding process that analysed similarities and differences among the cases to uncover contextual factors influencing the design and stewardship of PII. These factors include laws and regulations, the types of data-supplying actors, the nature of data exchange, and the presence of existing systems and standards. In the following sections, we will elaborate on these four contextual factors.

4.3.1 LAW AND REGULATION

Legislation and regulations are important contextual factors influencing stewardship and configuration. Usually, data sharing happens because of legal obligations. These obligations require both private and public parties to provide data and ensure its quality.

Additionally, regulations such as the GDPR, the Data Act, and government policy in the form of a comply-and-explain list often include rules governing the exchange of data itself. Although many laws and regulations are drafted at the national level, substantial regulations are also developed at the European level.

Cases firmly rooted in legislation do not always permit active development beyond that legislation (e.g., across chains). Furthermore, in certain instances, there is little interest in collaboration and development outside the PII boundaries. In such cases, we mainly observe single-actor stewardship. On the other hand, when legislation and regulations are less central, cooperation largely depends on the initiative of those involved. Here, shared stewardship greatly influences the structure of the PII.

The benefits of data sharing based on legislation are that they can often be implemented more quickly and are easier to enforce. Without the possibility of enforcement, incentives are needed to encourage participation. These include the fear of exclusion from decision-making processes, which could result in higher future investment costs for a company, thereby promoting a clear business case and supporting public interest and public value.

4.3.2 TYPE OF DATA SUPPLYING ACTORS

Another contextual factor is the type of organisations involved in the data exchange, particularly the supplying party. Here, we observe a difference between business-to-government and government-to-government data exchange.

In government-to-government information exchange, the influence of the public sector is strong, and the participation of public organisations is low. Cases within this category include pension funds, wage return statistics, BRP issue reporting, and property value notification. In this type of data exchange, private sector involvement is minimal, if any, typically involving private parties contracted by a public organisation through tenders. However, there is potential for multiple public parties to collaborate on shared systems, data formats, or standards, and to utilise public sector intermediaries. While shared systems can enhance

efficiency and standardisation, they also demand careful coordination. More often than not, stewardship in this context falls to a single actor or is partially distributed when outsourcing to a public intermediary. Nevertheless, the allocation of responsibilities remains limited, as government-to-government communication can continue to use government standards without conforming to market requirements. On the other hand, there are the business-to-government data exchange infrastructures. Examples of these cases include WOCO, SBR, eInvoicing, import declaration, and WMO. These involve (semi-)private companies. On the shared side, we observe that central systems are established through cooperation between different government parties. Cooperation can be pursued when data requests and processes are similar. At the front end, various private parties, on the other hand, offer software solutions and services to support these entities in their delivery. Not surprisingly, within these sectors, a significant number of private parties offer support services for arranging data exchange. Compared to other data exchange types, the private sector is strongly predominant. For example, within the case of eInvoicing, invoicing governments form only a small part of the invoicing process. This requires alignment between private service providers and government systems. Due to the diversity of systems, we are observing a greater distribution of responsibilities.

4.3.3 TYPE OF DATA EXCHANGE

The third contextual factor is the type of data exchange, which influences stewardship roles and configuration possibilities. On the one hand, we can identify people-centred data sharing. Examples include VUM, child welfare, vaccination, WMO, Care Codex, and KIK-V. In these cases, it is clear that this data sharing must be organised consistently at a national level. While this is acknowledged as challenging, effective regional processes and initiatives can be scaled up to the national level and utilised for multiple data-sharing projects. Many different actors participate in PII, and they must collaborate for the public interest or the interests of citizens. Because the impact is significant, cooperation is essential, and coordination is important. This also indicates a notable shift towards shared stewardship and consensus-building.

Conversely, we can identify trade- and Europe-focused data. Examples include migration, eCodex, imports, banking supervision, eInvoicing, and the Single Window. These areas transcend borders and are more globally oriented. The aim in such cases is often to reduce administrative burdens and standardise data exchange processes. Additionally, it shows that a greater level of European standardisation exists within certain sectors like trade, transport, justice, and migration. Several European working groups and standardisation committees operate within these sectors. European harmonisation is common in this context, and there is often flexibility in data formats. Since implementing new systems demands significant financial investment, and constantly changing data formats incur high costs, the development pace tends to be slow. Nonetheless, flexibility also introduces the challenge of supporting and adapting data formats to interpret them into internal systems. Within this context, PII stewards face the complex task of balancing stability, flexibility, and progress in PII development.

In several cases, we observe that data exchange is only partially coordinated at the European level. For example, in the cases of OOTS and eCodex, the focus is on technical standards for data exchange, while in eInvoicing, Bank supervision, and Migration, the primary emphasis is on the data content. However, in the case of SW, both technical and content-related aspects now fall within the scope of the European directive, with full harmonisation as the ultimate goal.

Each of these cases involves a European working group with Dutch representation. In some instances, only public stakeholders directly involved participate in the decision-making process. However, in cases such as eInvoicing, private stakeholders are also allowed to join the European working groups.

Active participation in decision-making at the European level is essential, as the resulting directives ultimately shape the organisation of the (national) information infrastructures. A national central working group functioning as a subcommittee of a broader international working group can promote balanced decision-making by incorporating diverse perspectives, from the public and private sectors to large and small organisations, thereby ensuring that the interests of all stakeholders are considered.

In the future, there will probably be a greater level of harmonisation at international levels. Depending on the domain, such as in trade and transport, this is expected to include not only Europe but also the broader global context.

4.3.4 PRESENCE OF EXISTING SYSTEMS AND STANDARDS

The final contextual factor we identified is the existence of established systems and standards. Often, the exchange of PII data involves converting paper information into a digital format. In such cases, existing data models, like the WCO data model used in trade and transport or the UBL for procurement messages, can be utilised. These models help improve the interoperability of data sets and enhance overall understanding of the data.

Furthermore, for the technical configurations in the PII, there are often existing systems and solutions available from the market or government. These include Digipoort, the Peppol network, exchange protocols like AS4, eIDAS, eDelivery building blocks, interface standards such as API, EBMS, and WUS, and security standards like the BIO. The government follows a comply-or-explain policy, requiring as many standards as possible to be used when establishing data sharing.

Such standards are developed at different levels and are sometimes formalised through trust frameworks. While these initiatives are often publicly initiated, we are seeing a trend towards evolving into public-private partnerships in order to broaden support and encourage wider adoption in the field.

Depending on the specific standards chosen, certain governance structures are linked to them. We observe that existing governance structures are in place around current systems and standards. The level of reliance on the standard will influence how the alignment with this governance is organised within the PII.

4.4 CHAPTER CONCLUSION

In this chapter, we have examined several examples of PII within the Dutch context. The description of the three exploratory cases offers an initial understanding of the differences in design choices in a PII. These three cases are used to validate the literature-based design variables discussed in the next chapter, which will develop a framework of design variables from both literature and practice.

Secondly, we have twenty-two cases that provide further insight into different types of PII configurations. Additionally, in these twenty-two cases, differences in the types of PII stewardship can be observed. Based on both literature and empirical evidence, three distinct stewardship types can be identified: single actor stewardship, distributed stewardship, and shared stewardship. A more detailed exploration of these types is presented in Chapter Seven.

A key finding in this chapter is that the design of PII and PII stewardship depends on cross-case contextual factors. These factors include law and regulation, the type of data-supplying actor, the nature of data exchange, and the presence of existing systems and standards. While contextual factors can influence design options and configurations, PII stewards can still make design decisions and guide the development of the PII in certain directions. The next chapter introduces research phase two *Design variable*, where the possible and relevant design variables are explored.

Chapter 5

DESIGN VARIABLES OF PII

This chapter presents the outcome of phase two: *Design variables*. As outlined in the preliminary chapters, PII stewards can guide the design of the PII by making design choices. These choices represent specific implementations of design variables. A design variable defines a dimension along which alternatives can be selected. We have considered the design choices that impact the overall design of a PII, including the information exchanged, technical components such as systems, standards, and services, and the governance structures involved. Additionally, we focus on the responsibilities of these elements, as they influence the distribution of responsibilities and thereby the PII stewardship. To gain a clearer understanding of the concept of PII and the range of decisions that PII stewards can make, this chapter will address the following sub-research question: **What are the relevant design variables of a public information infrastructure?**

We aim to explore the choices that PII stewards can make when managing a PII. To identify relevant design choices, it is necessary to examine the design variables more closely. Previous research has neither studied PII as a whole nor provided an overview of potential design choices. Each variable represents multiple design options, which may be selected by the designer, dictated by context, or influenced by prior decisions. It is important to recognise that a single design variable can include a range of design options.

The initial stage of this process involves conducting a comprehensive literature review of the design choices presented in existing research papers. This review offers preliminary insights into potential design variables. The review results in an extensive list of possible design variables within a PII. The next stage involves evaluating and refining the literature-based list by interviewing experts from the three PII cases: Custom declaration, Tax administration, and eInvoicing. To achieve this, we conducted seventeen interviews and analysed available documents for each case. This process results in a validated list of design variables and potential design options for each variable. Details on the research method of this part of the research are presented in Chapter 2.4.

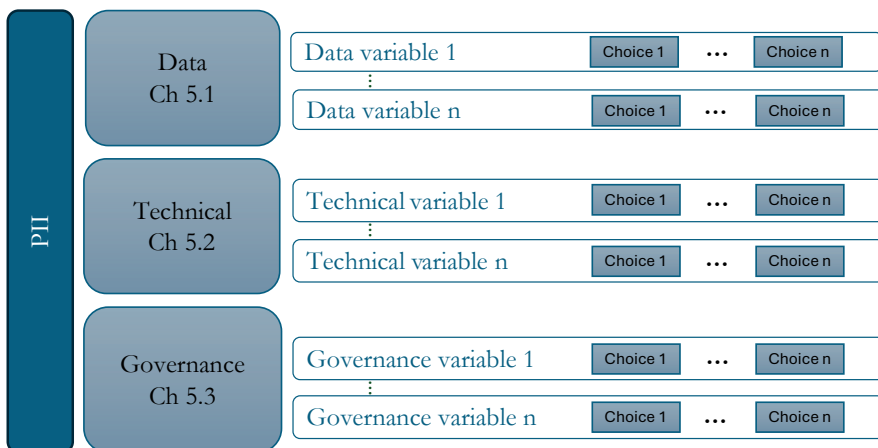


Figure 5-1. PII and the design variables

As presented in Figure 5-1, the analysis of the design choices is divided into three sections. Each subsection discusses the design choices for a component of a PII, including data, technical, and governance aspects. We begin with design choices in the data component, then proceed to technical design variables. The chapter concludes with an analysis of the governance design variables.

5.1 DATA DESIGN VARIABLES IN A PII

We begin this sub-chapter by examining design variables identified in the existing literature, which will then be validated and complemented through empirical research using case studies. In the last subsections of 5.1, we analyse the decisions made explicitly to enhance data quality.

5.1.1 LITERATURE-BASED DATA DESIGN VARIABLES

The main product in a PII is, of course, the information itself. Information can be structured through data formats, organised into meaningful datasets, and enriched with metadata to enhance its usability and quality. We conducted a systematic literature review, resulting in 27 usable papers. Based on the criteria described in Chapter 2.4.1, we identified seven data (format) design variables. These design variables are presented in literature-based data(format) design variables.

Table 11. Literature-based data(format) design variables

Design variables	Specification of possible choices	Authors
1. Level of information aggregation sent in PII	Aggregated Individual	Bharosa et al. (2011), Robinson et al. (2019).
2. Complexity of information sharing formats	Business rules, syntax, semantics, and technical	Lee et al. (2009), De Winne et al. (2011), He et al. (2012), Luciano et al. (2017), van der Kaa et al. (2018) Hofman (2019),
3. Level of accessibility of the information sharing formats	Open, closed	Bharosa et al. (2013), Hofman (2019),
4. Reusability of the information models	Standardised, specialised/custom	Lee et al. (2009), de Winne et al. (2011), Luciano et al. (2017), van der Kaa et al. (2018), Kurnia (2019), Hofman (2019).
5. Number of information sharing formats	One, many	Bharosa et al. (2013), van der Kaa et al. (2018), Hofman (2019),
6. How the use of the data formats is promoted	Power, financial, persuasion, policy, knowledge exchange, governance	Bharosa et al. (2018)
7. How the formats are managed	Central, decentral	de Winne et al. (2011), Bharosa et al. (2011)

The first design variable, driven by the literature, is the level of information aggregation. Some organisations prefer non-aggregated information, while others exchange aggregated data, such as over time, processes, or organisations (Bharosa et al., 2011; Robinson et al., 2019). There is a link between the type of

information and its level. Periodical information is often aggregated over time, whereas real-time information is usually individual data (Bharosa et al., 2011).

Unstructured digital data cannot be processed by receiving organisations, making data formats an important aspect of PII. A data format is a standardised structure used to exchange digital information in a way that can be interpreted by software or systems. Multiple authors discuss data formats, where design variables can be found (e.g., Bharosa et al., 2013; De Winne et al., 2011; Hofman, 2019; Kurnia et al., 2019; van de Kaa et al., 2018). These formats include variables related to complexity, accessibility, and reusability, as well as the number of data formats used, how the format is promoted, and how it is managed.

The second design variable concerns the use of a technical data format that describes syntax and semantics, thereby enhancing the interoperability of information sets (Lee et al., 2009; He et al., 2012; Hofman, 2019). Furthermore, complex formats often incorporate business rules between data elements. An example frequently cited as a complex data format is the XBRL format, along with its accompanying Taxonomy (de Winne et al., 2011; van de Kaa et al., 2018). The taxonomy serves as a dictionary of semantic data elements and can include rules (De Winne et al., 2011). Using an ontology to structure data, such as a taxonomy and other semantic representations, facilitates standardised reporting and enhances interoperability in the domain (Luciano et al., 2017).

A third and fourth variable relate to the accessibility and reusability of the data formats. Many formats are open and suitable for various types of information exchange. Often, authors promote the use of open formats to facilitate standardised, uniform information exchange (Lee et al., 2009; Bharosa et al., 2013). The standardisation of data might enable immediate traceability and easier comparison between data sets. (Luciano, 2017). However, in addition to standardisation, governments can also develop specialised data formats dedicated to information exchange.

Next, research by Van de Kaa et al. (2018) shows that multiple reporting formats are a common phenomenon. It can offer flexibility for organisational needs and stakeholder preferences. Depending on the context, it can be a challenge to harmonise data formats. Conversely, it also creates difficulties, as it may hinder consistent interpretation and increase the complexity of processing.

The sixth design variable, according to the literature, concerns how to stimulate the uptake of the format. To facilitate the adoption of a data format, Bharosa et al. (2018) propose six design options. First, the authority of a dominant organisation, which can act as a powerful catalyst. They might enforce data formats during data exchange. Second, financial incentives such as subsidies or cost-sharing mechanisms. These incentives can lower barriers to adopting a data format and make the transition more economically feasible. Third, persuasion can help shape perceptions, address resistance, and build support among stakeholders. Fourth, policy interventions, whether voluntary guidelines or mandatory regulations, can formalise expectations and stimulate the adoption. Fifth, creating a standard knowledge base that underpins the data format helps improve understanding among stakeholders. Finally, establishing public-private governance structures can ensure collaboration and align interests. Collectively, these instruments offer a comprehensive approach to steering the adoption process (Bharosa et al., 2018).

The final literature-based design variable identified concerns the management of data formats (or taxonomies). In some cases, a national standard may be used, whereas in others, a specific standard is employed with decentralised management. Here, certain stakeholders may be responsible for defining and managing the format (Bharosa et al., 2011). Examples of stakeholders include receivers and intermediaries. Together, these are the design variables for the level of information found in the literature.

This is the first time that design variables related to data (format) within a PII have been examined collectively. Therefore, further research is needed to determine whether all relevant variables have been properly identified. Additionally, it is essential to validate these variables and to refine or expand them as needed. Moreover, this empirical validation enables us to evaluate the practical relevance, consistency, and applicability of these variables in real-world settings. In the following subsections, we validate the data (format) design variables identified in the literature through our case study research.

5.1.2 VALIDATION OF LITERATURE-BASED VARIABLES AND EMPIRICAL DATA DESIGN VARIABLES

The respondents confirmed most of the variables identified in the existing literature. For instance, respondents across all three cases confirmed that a PII format is used to exchange data from the data provider to the receiver. The interviews highlighted the importance of standardised data formats. Unstructured data presents substantial challenges for receiving organisations, and data must be formatted to allow automated processing by the recipient. Formats can be semantic, syntactic, technical, or a combination of these. One benefit of structuring data is that it enables automated data validation during the exchange. With structured data, it is possible to validate for completeness (syntax), as well as accuracy and consistency (semantics). The choice may involve adopting a single format or multiple formats.

Nevertheless, three variables identified in the existing literature were not classified as design variables by the respondents. The first variable is the type of information exchange. The availability of real-time, individual, or aggregated periodical data was not regarded as a design element. The respondents indicated that the nature of the data was not considered a variable, but rather a context-dependent factor. Moreover, the openness of the formats used was not a standalone design variable but a combined design choice with standardisation of the data format and collaboration on it. Respondents noted that the data format is accessible to all who need it for data exchange and is often based on pre-existing data models, thereby ensuring some standardisation. In every case, the data format was characterised by openness and sharing. The literature also suggests that promoting the data format can be a design variable. However, it was believed that this does not directly affect the design of the PII. When promoting a format, one or more formats are already chosen and incorporated into the PII. Promoting a format is equivalent to promoting the entire PII, as it facilitates adoption and use; it can only indirectly influence the shape of the PII. Consequently, it is not considered a design variable.

Apart from the three variables that were not designated as design variables, three new variables were identified as design variables in the PII configuration. The first relates to collaboration on data formats.

The second concerns whether data formats are reused by or sourced from other organisations. This process of borrowing and integrating established formats helps ensure consistency across different PII and promotes the adoption of best practices. Finally, respondents noted various choices involved in designing and managing the data format. The following paragraphs further elaborate on these newly identified design considerations.

The first new design variable concerns whether there is collaboration on the data format. In all three cases, collaboration on the data format occurs; however, the degree of collaboration varies. In the SBR case, collaboration focuses on the use of taxonomies, mainly concerning semantic definitions. Because of XBRL, collaboration is also needed at a technical level. As XBRL is quite a complex technical standard, public organisations collaborate to share knowledge about the format.

In eInvoicing, central government organisations utilise a collaborative format. Multiple formats exist, and each involves collaboration. On the public side, there is cooperation on using a standard receiving format; this involves both syntax and semantics. *“The BFR describes what the governments need to process and receive invoices. It details the business rules translated into technical rules. We need these technical rules to validate invoices for compliance”* (R 2.1). Additionally, the government engages in international collaboration to align with European Union standards. At the national level, it works with private organisations to harmonise prevailing market formats. These collaborative efforts primarily focus on syntactic standardisation, with some attention to semantic alignment. However, due to the widespread use of XML in data exchange, no further collaboration on technical aspects is typically required. Finally, in the case of Customs Declaration, information sent through the PII is transmitted to multiple recipients. Since multiple recipients receive (parts of) the same message format, collaboration on the format is necessary between the recipients, covering both syntax and semantics. The Single Window system is designed to modify the technical aspects of the format. XML can be converted to EDIFACT and vice versa, depending on what is provided and the preferences of the receiving organisations. Based on the interview and case insights, we conclude that collaboration is a relevant design variable.

A second design choice mentioned by the respondents, which is absent from the current literature, involves reusing an existing data format. One of the main reasons for reusing data formats is to align with existing information exchange standards. As one respondent 2.17 stated, *“For the formats and interfaces, we decided to join what was already known to the market. (...) The choices were made to unburden the delivering parties as much as possible. (...) especially to limit the costs of market parties in their investments.”*

One of the main consequences of reusing existing data formats is that the organisation receiving the data loses control over the format's design and ongoing maintenance. Instead, this control may reside with external actors, which can include both public and private organisations operating at national or international levels. The design and development of the format might result from collaborative processes based on consensus-building, but it can also be influenced by a few actors with the authority to enforce their preferred format design. The lack of control can cause challenges such as reduced flexibility, slower responses to changing needs, or misalignment with internal processes. However, if reusing formats leads to

broader benefits such as global interoperability, improved harmonisation, or more efficient and reusable information exchange, the trade-off may be considered worthwhile.

The final new design variable concerns the design and management of the data format. In some cases, existing formats or parts of existing formats are used, meaning that format management is handled by external organisations. For example, in eInvoicing, private organisations manage formats such as Peppol BIS 3.0. A public service provider, Logius, manages these formats on behalf of receiving public organisations. In the Customs declaration, one of the receiving organisations, the tax authority, manages the format for all recipients. In the Tax administration case, the public service provider governs the data format process. The case studies show that designing and managing data formats usually involve a range of stakeholders. These may include data receivers, public or private intermediaries, data providers, and policymakers, each possibly playing a different role in shaping or overseeing the format. Stakeholder involvement varies depending on factors such as the type of data exchange, the nature of the data exchange processes, the political or business interests related to the format, and the complexity of the data format itself. As a result, the stakeholder configuration—and the potential for coordination issues—can vary greatly across different contexts. Therefore, we consider this a design variable.

Between Tax administration and eInvoicing, there is a difference in the ability to influence the data format. As respondent 2.9 stated: *“In SBR, the requesting organisation have the power to form the taxonomies and the responsibility for the choices they make.”* In eInvoicing, however, receiving organisations have less influence over the format and have to settle data format requests through a public governance structure. The government formats need to be agreed upon by all receiving organisations. In a Single Window system, the data provider submits information using a single standardised format. The Single Window then distributes this information into multiple data formats tailored for each receiving organisation. Consequently, any changes to the format by a receiving organisation affect the overall data format but do not impact the other receivers. This places the primary responsibility for implementing and adapting to the format on the intermediary distributing the information and the information provider. Respondent 2.13 explained why there is no harmonisation between the receiving organisations: *“Due to the governments’ calendars, it is not possible to arrange one format for receiving organisations.”* This results in the final overview of design choices on the level of information (formats) as presented in Table 12.

Table 12. Information design variables

Variable names	Design Options			
	Yes	No	Partly	
Formatted data	Semantic	Syntax	Technical	Combination
Complexity of information sharing formats	Collaboration	Individual	Both	
Data format collaboration	One	Multiple		
Number of information sharing formats (from the receivers' perspective)	Standardised	Custom	Both	
Wider use of the information models	Yes	No		
Reuse of format	Public	Private	Both	None
Responsible for creating a data format	Intermediate	Receiver	Third party	Sender
Responsible for managing data format	Public	Private	Both	None
Responsible for managing data format	Intermediate	Receiver	Third party	Sender

5.2 TECHNICAL DESIGN VARIABLES IN A PII

In this sub-chapter, we explore the technical design variables of a PII, including choices regarding systems, services, and standards in the PII. As in Chapter 5.1, we start by examining design variables identified in existing literature. These are validated and complemented by the case studies. Finally, we analyse the decisions that specifically contribute to managing and enhancing technical quality.

5.2.1 LITERATURE-BASED TECHNICAL DESIGN VARIABLES

Without systems, information cannot be digitally exchanged between actors. An important part of the PII is therefore the technical design variables, including the choices on systems, provided services, and supporting standards. We conducted a systematic literature review, resulting in 27 usable papers. Based on the criteria described in Chapter 2.4.1, we identified twelve technical design variables. These design variables are presented in Table 13.

Table 13. Literature-based technical design choices

Design variables	Specification of possible choices	Author
1. Type of process	Push-Pull	Bharosa et al. (2011), van Engelenburg et al. (2015), Fu et al. (2019), Rukanova et al. (2021), Van Engelenburg et al. (2018)
2. Reuse of information	Piggybacking (store once, use many)	Bharosa et al. (2013), Robinson et al. (2019)
3. Type of infrastructure	Central, semi-central, decentral Dyadic, Multilateral, or hybrid	He et al. (2012), Paide et al. (2018), Chen et al. (2019), Kurnia et al. (2019), Rukanova et al. (2021)
4. Use of standards	Focus on the use of standards, no focus on the use of standards.	Bharosa et al. (2013), Vest et al. (2014), Wendel et al. (2015), Robinson et al. (2019).
5. Interfaces	Web-based – SOAP	de Winne et al. (2011), Bharosa et al. (2011), He et al. (2012)
6. Encryption	Focus on encryption, no specific focus on encryption.	He et al. (2012), Engelenburg et al. (2015), Engelenburg et al. (2018), Chen et al. (2019)
7. Service provision	Platform or portal	De Winne et al. (2011), Bharosa et al. (2013), Paide et al. (2018), Robinson et al. (2019), Zeng et al. (2023)
8. Costs of PII	Pay for use, Central government funding.	Mermoud et al. (2019), Klievink et al. (2012), Bharosa et al. (2013)
9. Way to control use (adoption)	Incentives, regulation, governance	Bharosa et al. (2011), Bharosa et al. (2013), Klievink et al. (2014), Susha et al. (2019)
10. User friendliness	Focus on user friendliness, or no focus on user friendliness	Robinson et al. (2019), Wendel et al. (2015)
11. Service provision	Validation (compliance by design) Combining information sets Formatting information Time stamping Authentication, authorisation, security	He et al. (2012), Bharosa et al. (2013), Vest et al. (2014), Wendel et al. (2015), Engelenburg et al. (2018), Paide et al. (2018), Chen et al. (2019), Rukanova et al. (2021),
12. Public or private?	Contracted PPP Public Private	Bharosa et al. (2011), Vest et al. (2014), Wendel et al. (2015), Paide et al. (2018), Mermoud et al. (2019), Robinson et al. (2019), Rukanova et al. (2021)

The first two variables identified are the choice between a push or pull process and the reuse of the exchange process. In an information push process, businesses proactively submit information to governments, whereas in an information pull model, governments retrieve the information from the systems of data providers (Bharosa et al., 2011). This choice affects how a PII is configured, since an information pull necessitates decisions regarding information access (Bharosa et al., 2011; Fu & Zhu, 2019; Rukanova et al., 2021; van Engelenburg et al., 2015; Van Engelenburg et al., 2018) and storage (Kurnia et al., 2019; Rukanova et al., 2021). Information-pull infrastructures encounter challenges related to data access, especially necessitating authentication mechanisms for the data recipient. Conversely, in an information push model, the information provider must authenticate and identify themselves. Furthermore, information-pull infrastructures enable the reuse of datasets (piggybacking), as data providers can simply grant access to additional organisations without altering the data or the process itself. Through piggybacking, multiple organisations can access the same dataset, which is more common in a data pull scenario. As Bharosa et al. (2013) state, “existing data and data flows are reused flows for a goal other than that for which this data was initially intended.” Piggybacking is less frequent in information push models, where data must be actively

sent to each recipient, as this involves customised data exchange. Both businesses and governments may benefit from a pull mechanism (Robinson et al., 2019).

The third design variable is the type of infrastructure. There are categories such as central, semi-central, and decentral (Chen et al., 2019; He et al., 2012; Kurnia et al., 2019; Paide et al., 2018; Rukanova et al., 2021; Zeng et al., 2023), as well as dyadic-multilateral-hybrid (Kurnia et al., 2019). In a dyadic system, a connection must be established with each individual receiver, resembling a decentralised infrastructure. A multilateral system allows connections with multiple stakeholders without requiring a direct link between every receiving organisation. In this case, the information exchange is centralised within a single system. An intermediate option is the hybrid system, where a providing party can connect via both dyadic and multilateral linkages (Kurnia et al., 2019).

Many design choices in the literature appear to be made by individual organisations. However, some design variables are agreements or regulated across all systems, such as the use of standards or encryption, involving the fourth to the sixth design variables. Standardisation involves agreements or protocols that specify how multiple systems are used and interconnected (Bharosa et al., 2013). It can improve interoperability between systems and prevent organisations from repeatedly reinventing the wheel (Vest et al., 2014). Overall, the output of the information infrastructure can be maximised if systems are compatible (Wendel & Edberg, 2015). Compatibility across the entire infrastructure can be achieved through standardisation of interfaces, such as SOAP, and exchange protocols between systems (He et al., 2012; De Winne et al., 2011; Bharosa et al., 2011). Regarding standardisation for encryption, it is possible to send information through the infrastructure in encrypted form, so only the receiver decrypts it (van Engelenburg et al., 2015). Chen et al. (2013) highlight the importance of effective access control to documents and secure exchange through encryption. Encrypting data, however, can result in limited services provision during transmission, such as validation or combining information sets, since encrypted messages pose challenges for these functions.

A seventh variable from literature is the use of gateways and intermediating platforms in information exchange (De Winne et al., 2011; Paide et al., 2018; Robinson et al., 2019; Zeng et al., 2023). Gateways serve as portals through which information is transferred, and by adding services, they can develop into platforms (Zeng et al., 2023). These platforms are described as one-stop shops, Single Windows, or facilitating infrastructure providing common functionalities (Bharosa et al., 2013). They streamline business processes and can optimise service delivery, fostering collaboration among multiple actors (Zeng et al., 2023). Often, they offer multiple services, such as security, authentication, and information validation mechanisms. They provide a single point of contact for businesses. Governments can either enable other organisations, such as private businesses, to design and maintain these platforms or provide them directly (Zeng et al., 2023).

The following three variables relate to the funding of the systems, user-friendliness, and the method of controlling the information infrastructure. Research indicates that, in many cases, earnings derive from subscriptions and fee-per-message charges, particularly in systems operated by private service providers (Klievink et al., 2012). Data providers and recipients lacking the necessary skills to support data exchange often have high demands and a strong willingness to pay (Wendel & Edberg, 2015). In this context, user-

friendly solutions are developed to reduce the burden on both data providers and recipients during data exchange. When a viable business case exists, private organisations can benefit from offering larger solutions within a PII, with most costs borne by users. However, it is also possible for governments to provide financial incentives to encourage stakeholder participation in the PII (Mermoud et al., 2019). “Governments can play a role to stimulate data sharing by companies for social projects by putting in place relevant policies and incentives for companies” (Susha et al, 2019, p.233). Such incentives, including social or financial rewards, apply not only to infrastructure providers but also to users. Incentives may be necessary when organisational goals do not align with those of the information infrastructure. There are additional methods to promote the use or adoption of an information infrastructure, such as governance guidelines, laws, and regulations (Bharosa et al., 2011). Governance structures can be established to enable stakeholders to collaborate through formalised relationships (Klievink et al., 2014).

Second to last, the literature presents several possibilities for service provision during the information exchange process (e.g., Bharosa et al., 2013; Paide et al., 2018; Van Engelenburg et al., 2018; Vest et al., 2014; Wendel et al., 2015). The services offered within infrastructural components all transmit information through the information infrastructure in some manner. However, beyond merely transmitting information, literature discusses many potential services such as combining information, transforming it into a format required by the recipient (Wendel et al, 2015), timestamping (Paide et al., 2018), and trust services like authentication, authorisation, and security (Bharosa et al., 2013; Paide et al., 2018; Van Engelenburg et al., 2018; Vest et al., 2014). A prominent service frequently provided within the infrastructure is information validation. Literature on information infrastructures in reporting information also refers to this as compliance by design, which is built into the infrastructure itself. “Monitoring compliance requires data, sometimes in the form of template-based reports that serve either as evidence or declarations that the citizen or organisation complies with a business rule, guideline, or legal regulation” (Bharosa et al., 2013). The validation of information can be carried out by regulators, intermediaries, or software providers of the requesting organisation.

The final variable, implicitly indicated by the other variables, is the involvement of the private sector. Different sources of literature mention various cases involving both private and public organised components in the infrastructure (Bharosa et al., 2011; Rukanova et al., 2021; Vest et al., 2014; Wendel et al., 2015). There are various forms of collaboration, including public-private partnerships. Research shows that multiple types of stakeholders might be involved in the information sharing process; these include, but are not limited to, governmental organisations, technology providers or intermediaries, standardisation organisations, and businesses (Rukanova et al., 2021). Governments can serve as receiving organisations, with their own systems for collecting and processing information. However, there are also intermediary governmental organisations that provide services for the benefit of receiving organisations (Bharosa et al., 2013).

Together, these are the technical design variables for the level of systems, services, and standards found in the literature. This is the first time technical design variables within a PII have been examined collectively. Therefore, further research is needed to determine whether all relevant variables have been adequately

identified. Additionally, it is important to validate these variables and refine or expand them as necessary. Additionally, this empirical validation enables us to evaluate the practical relevance, consistency, and applicability of these variables in real-world settings. In the following subsections, we validate the technical design variables identified in the literature through our case study research.

5.2.2 VALIDATION OF LITERATURE-BASED VARIABLES AND NEW EMPIRICAL TECHNICAL DESIGN VARIABLES

The respondents confirm most of the variables identified in the existing literature. A major design decision in the PII concerns the type of infrastructure. This respondent distinguishes different options by combining the earlier two models into one: central, semi-central, decentralised, distributed, or hybrid models. Distributed information infrastructures involve a dyadic connection, where a connection needs to be made with each receiver. In a decentralised information infrastructure, there are multiple hubs between senders and receivers, which multiple organisations can use.

A second factor confirmed by the respondents is the use of existing systems. In some cases, collaboration involves integrating with current systems rather than creating new ones from scratch. Several respondents noted that the decision depends on system requirements and the availability of suitable existing systems. The choice to use existing systems or standards can be strategic, dictated by legacy constraints, or driven by legislative compliance.

Besides the use of existing standards, respondents emphasised their importance. Within the Dutch government, there is an ‘apply or explain’ list that encourages the reuse of standards and promotes uniformity across agencies. This means that there is a choice between leveraging what already exists, using (parts of) it, and exploring new techniques or creating a system from the ground up based on how it should be designed.

Next, the variable of push or pull was validated. Whether a PII operates through a push or pull model is partly influenced by the type of data it handles. Although all examined cases used a data push approach, respondents know that other types of PII might involve a data pull. This distinction fundamentally shapes the configuration of PII, as it changes the types of services and, to some extent, the responsibilities for quality along the chain. It is therefore not excluded that having a data push is a design variable. It was observed that data pull infrastructures commonly offer services for accessing data, whereas such services are less frequently found in data push infrastructures.

The last confirmed variable is the involvement of the public sector. In all three cases, to some extent, the private sector provides systems and services in the PII. In all cases, the market involved helped at the front of the data exchange process. Private organisations engage in software development, providing data management tools, acting as an access point to the network, providing additional data exchange services such as signing, or providing multiple.

Based on the interviews, we conclude that some variables are not part of the design framework. During the interviews, it became clear that the adoption of the PII was not primarily a design choice regarding the technical aspects but rather a variable related to the governance of the PII. Therefore, it is not included in the framework for design variables in Systems, Services, and Standards. Secondly, the costs of the PII depend on the systems and are not a decision made at the level of the PII itself. In all cases, private organisations provide services on behalf of the information-providing organisations. In these cases, payment is typically made for usage. For public organisations, different systems with various funding methods are standard. However, these depend on collaboration and stakeholders and are not solely a funding issue for the entire PII. Finally, the decision to install a platform is not considered a design choice. Having a platform results from several other choices, creating a central system with specific services. Therefore, this is not included within the framework of design choices in a PII.

Some of the choices were mentioned as part of another design variable; for example, transforming formats is just one of the services that can be provided. Additionally, during the interviews, it became clear that providing a portal, encryption services, or specific interfaces is part of the service provision in the PII and depends on the use of standards. Following the interviews, a comprehensive list of services was compiled: providing a submission service, an address book, supporting the delivery process, ensuring or validating information quality, data formatting, information transformation, identification and authentication, providing support, offering message track & trace, safety & security, and securing/storing information. Each of these services can be provided by one or more organisations within the PII, and each organisation can provide one or more of these services. As respondent 2.11 stated: *“We build services in a modular way, giving us flexibility. We started with a receiving and delivery service, an authorisation service, and a validation service. With the increasing number of message types, we added assurance, attachments, and signatures of the accountant. It is easy to incorporate new services.”* However, it is also evident that these services are not static and may (and will) evolve.

Additionally, respondents observed that through piggybacking, the PII remains unchanged, whereas the principle of “reporting once, used by many” would. During the interviews, specifically in the context of Customs declarations, respondents highlighted the advantages of a collaborative model in which data-reporting organisations submit their information only once. The infrastructure then ensures this data is shared with all relevant stakeholders, thereby reducing administrative burdens.

Based on the interviews, we have identified three new variables that should be introduced. The first variable introduces the relationships between PII users. For instance, in Tax administration, multiple data providers send data to a single organisation, the Tax Administration. In Customs declarations, a data provider can send the same data to multiple recipients. For eInvoicing systems, there are many providers and recipients on both sides of the PII. These observations lead to four possible scenarios in PII: 1-1, 1-N, N-1, and N-N.

The second is the choice to collaborate in designing, using, and maintaining systems. In all three cases, collaboration is considered a crucial aspect of PII design. In each case, there is collaboration from the public side. Multiple information exchange processes are so aligned that a large portion of the technical

components of the PII are reused. For example, Tax administration and eInvoicing share the Digipoort system. However, more information chains also use Digipoort. Digipoort is a collaboration among public organisations, managed by Logius, an intermediary organisation providing infrastructure on behalf of the collaborating entities. Respondents 2.14, 2.15, and 2.16 note that responsibility for this part of the process is transferred to Logius. The Single Window is another excellent example of public collaboration, where multiple public organisations work together to establish a single portal for various message types in Customs declarations. RWS, NVWA, BD, and KMar joined forces, although BD is responsible for executing the systems. In procurement messages via Peppol, the organisation OpenPeppol determines the network, standards, and governance. OpenPeppol is a collaboration between public and private organisations that jointly set the network requirements.

When organisations collaborate, assigning responsibility may be more difficult. By appointing a single intermediary to oversee the execution of the system or standards, this issue can be mitigated. Collaboration is often coupled with a governance structure, which manages decision-making, communication, and execution. More about governance structures is explained in chapter 5.3.2. Not all respondents plead for collaboration. Respondent 2.3 remarks that many collaborations lack clear responsibility, which can lead to unsuccessful PII. Additionally, cooperation with the private sector is not always actively pursued. As respondent 2.13 observes: *“a vision of the government is that we are not involved with the private software developers. They are able to provide more efficiency for information supplying organisations. These developers can integrate their information exchange services with other external services outside of the government. For example, a data management system and an information exchange service. We should leave those services to the private sector.”*

The third variable, which we explicitly include in the framework, is the roles responsible for providing the systems and services. Several roles are often involved, such as the private information provider, the private intermediary, the public intermediary, the receiver, and third parties. In the providing organisations, interviewees distinguished between large and small organisations. Many large organisations did not outsource services such as submitting and formatting. A respondent 2.17 in Customs declaration stated: *“In the case of declaration through MSW, large companies can create their own submitting service and request an S2S connection with MSW. For smaller businesses, or businesses that want to outsource this service, the market offers Software and H2S portals.”* Besides size, complexity influences the decision to outsource. XBRL is a complex language for providing information, and private intermediaries offer portals or solutions to overcome this XBRL challenge and simplify the process for information providers.

In the case of Customs declarations, one of the Receiving organisations provides a system for other organisations, thereby acting as an intermediary organisation. In addition to validation, no other services in the information-sharing process are reported to be provided by any of the receiving organisations.

In the other two cases, a public intermediary provides a significant infrastructure component that includes services such as validation, format transformation, and submission portals. In the cases of eInvoicing and Tax administration, a public intermediary called Logius is responsible for the information-sharing process on behalf of many receiving organisations. Logius offers validation services, identification and authentication services, and a status information service. For Tax administration, Logius also has an

XBRL knowledge centre that helps organisations design data formats. For eInvoicing, Logius manages all government data formats and provides implementation support for receiving organisations. It also manages the OIN registry. In the case of SBR, there is a greater need for identification, since respondents mentioned that submitting this information is a legal obligation, so they need to ensure they are providing the correct information. Information exchange uses PKIo certificates, which act as digital passports for additional identification. This registry is maintained by the public organisation named Logius.

In the case of eInvoicing, private intermediaries play a large role in submission, information transformation, and data formatting. Here, the private intermediary acts as the connection between the accounting systems and the financial dealings with customers and suppliers. The government is not the only entity that businesses report to. Private intermediaries help organisations to procure from many different buyers, each with their own specifications. There are also methods to send procurement messages without intermediaries, but these are only used when dealing with the government. A respondent 2.1 stated: *“There is a law, which says that if the market can provide a business, the government cannot disturb that market. This is why the scope of the portal is only for the central government. The scope of market portals is to connect to the entire world.”*

In the case of Tax administration, the XBRL format is quite complex, creating a business case for private intermediaries to assist organisations in formatting information. Private organisations offer portals or solutions to overcome this XBRL barrier and simplify the process for the information provider. The complexity of XBRL does limit the market, and fewer organisations provide software for SBR than for eInvoicing, which uses XML.

In the case of the custom declarations, respondent 2.13 stated that for declarations made through MSW, large companies can establish their own submission service by requesting an S2S connection with MSW. For smaller businesses or those wishing to outsource this service, the market offers Software and H2S portals. Intermediaries mainly provide methods for submission and data formatting services.

Besides public and private intermediaries, there are also third parties involved in some of the cases.

In SBR, the quality of information is especially important to the receiving organisation. This has led to accountants being involved in the process. Their role is to approve the quality of the information, ensuring that what is reported reflects reality. None of the other cases included this type of validation beforehand. In the other two cases, eInvoicing and SBR, the information is validated by public intermediary systems. However, this validation only ensures that the format of the information is correct, allowing the recipient to process the information set.

Other third parties involved in the information infrastructure for eInvoicing include the Netherlands Enterprise Agency (RVO), which assists organisations with their B2G eInvoicing processes with the government. When significant errors occur, they contact the businesses or vice versa for assistance.

Next are the Dutch Peppol authority and the Telecommunications authority. The Dutch Peppol authority is part of the Open Peppol initiative. In the Netherlands, the Peppol authority is public, but OpenPeppol is a private organisation. Besides managing the address book, the Dutch Peppol authority also oversees the service providers, which can be access points to the Peppol network. To become a service provider, a contract must be established with the Dutch Peppol Authority. The Dutch Peppol Authority

sets the rules and regulations that access points must comply with. They establish policies for the future safety of the Peppol network. The telecommunications authority regulates the use of Peppol in the Netherlands and enforces compliance among organisations.

The combination of empirical interviews and literature results in the framework of technical variables, as presented in Table 14.

Table 14. Technical design variables in a PII

<i>Variable name</i>	<i>Design Options</i>				
<i>II type</i>	Central	Semi-central	Decentral	Distributed	Hybrid
<i>Infrastructural relation</i>	1-1	1-N	N-1	N-N	
<i>Type of process</i>	Push	Pull			
<i>Report once, used many.</i>	Yes	No			
<i>Reuse of existing systems</i>	Yes, nationally	Yes, international	No		
<i>Use of standards</i>	Yes	No			
<i>Collaboration in the PII</i>	Individual	Collective (private)	Collective (public)	Collective (both)	
<i>Provided services</i>	Providing a submission service, maintaining an address book, supporting the delivery process, ensuring or validating information quality, data formatting, transforming information, performing identification and authentication, offering support, supplying track & trace of messages, ensuring safety & security, securing and storing information, implementing encryption, and enhancing user friendliness.				
<i>One or multiple services</i>	One	Multiple			
<i>Involved in providing technical components</i>	Sender	Intermediate	Receiver	Third party	
<i>Involved in providing technical components</i>	Public	Private	Both		

5.3 GOVERNANCE DESIGN VARIABLES IN A PII

In this third sub-Chapter, we examine the governance design variables in the PII. As with Chapters 5.1 and 5.2, we start by reviewing the design variables identified in existing literature. These are validated and supplemented by an empirical case study. Finally, we analyse the decisions that specifically contribute to managing and improving governance quality.

5.3.1 LITERATURE-BASED GOVERNANCE DESIGN VARIABLES

Actors play a crucial role in a PII, since infrastructures are inherently socio-technical systems. They not only influence and shape the design and functionality of the PII, but also affect how it is used and maintained over time. Furthermore, the PII governance configuration can be organised in various ways, with one or multiple governance structures. Each governance structure in turn shapes the roles and responsibilities, highlighting the interdependence between social and technical elements within the infrastructure. We conducted a systematic literature review, resulting in 27 usable papers. Based on the criteria described in Chapter 2.4.1, we identified six governance design variables. These design variables are presented in Table 15.

Table 15. Literature-based governance design variables

Design variables	Specification of possible options	Authors
1. Scope of governance structures		Kurnia et al. (2019), Bharosa et al. (2013),
2. Presence of multiple levels in governance configuration	No, yes	Bharosa et al. (2013)
3. Organisations included in the governance structures.	Public, private & government, information providers, intermediaries, and software providers.	Bharosa et al. (2013), Mermoud et al. (2019), Robinson (2019), Rukanova et al. (2020), Susha et al. (2019), Vest et al. (2014), Wendel et al. (2015)
4. Ordering of the governance structures in the configuration	Hierarchical, network, heterarchical, market	Kurnia et al. (2019), Klievink et al. (2012), Mermoud et al. (2019)
5. The type of control mechanisms	Traditional command and control mechanism- relationship management. Contract vs relational governance	Aben et al. (2021), Rukanova (2021), Klievink et al. (2014), Klievink et al. (2012)
6. Funding of the governance structures	Funded by government subsidies or private funding.	Bharosa et al. (2013), Vest et al. (2014), Mermoud et al. (2019)

The first design variable identified in the literature we found is the scope of governance structures. According to Kurnia (2019), governance structures must address information ownership, decision-making, and the mechanisms for controlling decisions. Governance scopes can be, for example, to reduce administrative burden and promote the re-use of information (Bharosa et al., 2018), but also “ensure that conflicts are addressed, and that the necessary resources are properly allocated and utilised” (Rukanova et al., 2020).

The second variable relates to the number of governance structures and whether a single structure contains multiple levels. The overall governance configuration can consist of a single governance structure or include multiple governance structures, and each of these governance structures can include multiple levels. Examples of these levels include but are not limited to decision-making boards, coordinating platforms, and expert groups (Bharosa et al., 2018).

Regarding the third variable, several authors mention different types of stakeholders in governance (Bharosa et al., 2018; Mermoud et al., 2019; Rukanova et al., 2020; Robinson et al., 2018; Susha et al., 2019; Vest et al., 2014; Wendel et al., 2015). It becomes evident that, in a public II governance configuration and its structures, both public and private organisations can be involved. Besides the distinction between public and private, the literature indicates that multiple actor groups can be involved or represented in governance. These include, for example, information providers, intermediaries, and software providers (Bharosa et al., 2018).

In cases with multiple governance structures forming the governance configuration, the literature identifies a fourth variable. The coordination between governance structures is reported to have three orderings: Hierarchical, Market, and Network. Hierarchical orderings are characterised by a top-down or bottom-up mechanism (Klievink et al., 2012; Kurnia et al., 2019). Market orderings are governed without a hierarchical relationship between them (Klievink et al., 2012), and networks are positioned between markets and hierarchies. Networks tend to rely more on informal practices based on shared interests (Klievink et al.,

2012). Kurnia (2019) identifies an additional heterarchical ordering, situated between hierarchical and network ordering, as a combination of both.

The fifth variable, derived from the literature, represents the type of control mechanisms employed by governance structures. Depending on the type of information infrastructure and the stakeholders involved, a different governance configuration might be selected. In a free-market model, the government functions more as a facilitator than a driving force in governance (Mermoud et al., 2019). Particularly in multi-actor contexts, there are governance situations where governments lack formal power (Rukanova et al., 2020). Traditional power often exists in hierarchical governance structures and is frequently reinforced through contractual governance. (Aben et al, 2021). Whereas governance networks are often more open to distributing power among different authorities, both public and private (Klievink et al., 2012). These types of governance configurations depend more on trust between businesses and governments, based on common or societal interests (Aben et al., 2021; Bharosa et al., 2013).

The final variable from the literature concerns the funding of governance and overhead costs. We observe that government-oriented governance structures are more frequently funded through government subsidies, whereas public-private governance is typically financed through individual contributions for participation (Mermoud et al., 2019).

These are the governance design variables identified in the literature. This is the first time that technical design variables within a PII have been examined collectively. Therefore, further research is required to determine whether all relevant variables have been accurately identified. Moreover, it is essential to validate these variables and refine or expand them as needed. Additionally, this empirical validation enables us to evaluate the practical relevance, consistency, and applicability of these variables in a real-world context. In the following subsections, we will validate the governance design variables identified in the literature through our case study research.

5.3.2 VALIDATION OF LITERATURE-BASED VARIABLES AND NEW EMPIRICAL GOVERNANCE DESIGN VARIABLES

The respondents confirmed the majority of the variables identified in the existing literature. During the interviews, it was reaffirmed that governance structures involve the distribution of responsibilities between the organisations involved. Across the various cases, different governance scopes exist, resulting in one or more governance structures. During the interviews, we established how the ordering of the governance structures can appear in the governance configuration, as presented in Figure 5-2

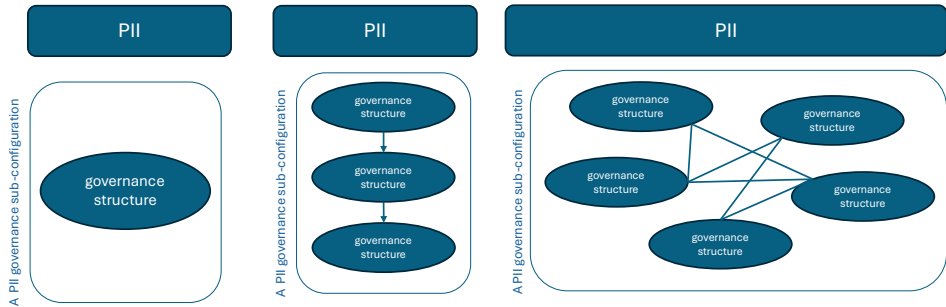


Figure 5-2. Different governance structures

“A *governance structure* has formal goals, ownership, distribution of property rights, and decision-making procedure, that fosters and sustains collective action” (Enjolras, 2009). The government implemented some of the governance structures; others already existed. Over time, structures can be created and cancelled as needed. However, they are installed through a legal foundation, as a governance structure refers to the formal distribution of rights and responsibilities among the different participants (Brooks, 2004), providing means by which (parts of) the PII is managed as an entity. Each governance structure can have a distinct scope, including different responsibilities. Many have representatives in other structures, resulting in a large network of governance structures. The complexity of multiple governance structures, according to respondent 2.5, lies in responsibility and funding: “Everyone believes progress needs to be made, everyone sees it as development; however, who is in charge? (...) I sometimes have no clue who is working on which topic”. The number of governance structures appears to be related to the scope of governance, the participants involved, and the types of structures. Having a single central governance structure can increase the number of stakeholders involved in the governance structure, making alignment more difficult. However, decentralised governance structures can lead to fuzzy decision-making, unclear responsibilities and alignment issues.

Respondents identified two orderings applicable to the public sector if there are multiple governance structures in the PII: the hierarchical and network orderings of governance structures. In the cases of both Tax administration and eInvoicing, multiple governance structures exist. In the case of Tax declaration through SBR, the governance structures address different topics that complement each other. There is a clear definition of each governance structure with its own responsibilities and hierarchical structures. In the field of eInvoicing, PII governance configuration represents more of a network ordering. Many of the structures coexist, some with and some without a formal hierarchy. Formats are governed at both national and international levels. Some, such as the national governance structures, involve only public organisations, while others, like OpenPeppol, include both public and private organisations. Respondent 2.4 noted: “Each of the governance has a different purpose. CEN is for creating a format, not the use or implementation of the format. Peppol has the purpose of implementing the use of Procurement information exchange. Nationally, there are governance structures to align market and government in a national format.” It is noticed that some authorities overlap in parts of the PII, creating fuzzy decision-making rights.

We identified that the presence of multiple levels across the cases is very similar. In all cases, governments have multiple levels for decision-making and preparation. Regarding Customs declarations, decisions are being prepared at several layers, both on financial and technical issues. According to a respondent, there are four levels: operational workgroups, receivers' assembly, change advisory board, and the strategic or control committee. For Tax administration in both governments, there is a council for strategic decision-making, platforms for tactical decisions, and multiple expert and working groups for signalling, advising, and preparing decisions. In the case of eInvoicing, there are many governance structures of various shapes and sizes, though we recognise the same structure of preparation through expert or working groups and higher decision-making committees. There are some downsides to having multiple layers. As a respondent, 2.13 stated, "*Some of the discussions in the work group will be dragged towards the higher levels, and then discussions will run across. It is then not clear who can make which decision and when. (...) At a higher level of governance, you would like to say, 'we're doing this. No discussion.', but then there is critique from lower levels.*" Within the case of Tax administration, some respondents state that the multiple levels have become too complex and therefore a reorientation of the governance is needed.

In the governance of SW, all involved public organisations in the case were represented. In the governance of SW, it was deliberately decided not to include the private sector of the information exchange network. Respondent 2.13 mentioned that there are many dialogues with the private sector, but they have no decision-making rights and are therefore not included. According to the respondents, involving the private sector in governance added another layer of complexity due to differing values and interests. Alongside the official governance structure, many recipient organisations are in dialogue with private organisations or their branch representatives. Respondent 2.13 stated, "*There have been large improvements in collaboration between the governmental organisations. Being unknown also made being unloved. In the case of Customs declarations, the governance participants needed time to learn how to collaborate, understand each other, and recognise each other's interests. This needed time and has led to excellent cooperation.*" Respondent 2.14 stated that collaboration is very high in the case of Customs declarations. Involved stakeholders may therefore differ between different PII configurations.

In the case of eInvoicing, the involvement of private organisations varies among different governance structures. There is a governance structure involving only the public sector, including public intermediaries and public receivers. However, there are also governance structures involving private organisations. Peppol, for example, is part of the PII with its own governance, encompassing all parties using Peppol. Part of the B2G information exchange utilises the Peppol data format and infrastructure. The government, represented by the Dutch Peppol authority and Logius, is part of this governance but has limited or no control over the decisions made. Involving more types of stakeholders in governance can reduce the control each stakeholder has if there is a consensus-based decision-making process. Given the different goals and values among the stakeholders, reaching consensus can be challenging. Nonetheless, once consensus is achieved, higher adoption can be attained, enabling rapid development of the PII.

The design variable for enforcing decision-making splits into two variables. The first is the decision-making rights within a governance structure, and the second is how these decisions are enforced. The

decision-making rights are also highly dependent on inclusiveness and the stakeholders involved in governance. In the governance of customs declaration through the Single Window, consensus is sought. At lower levels of governance, decision-making is prepared and coordinated. The final decision is then made at the highest level by the control committee. If lower levels do not agree, the issue can also be escalated to the control committee, in the hope that they will reach a mutually agreeable solution. As respondent 2.13 stated, “On the public side, you try to have consensus. If one of the participating organisations really does not want something, it becomes difficult. It is not the case that the majority rules. In practice, it helps if governmental organisations understand each other’s interests and realise they are working together on a common goal. Sometimes conflicting interests can be changed by looking at different ways of financing decisions.” Due to this consensus-based decision-making process, effective communication is of great importance.

Decision-making in the governance structure of SBR is clearly outlined in the governance structure documented by Logius. However, the government retains most of the decision-making rights. Respondent 2.12 noted that, due to the importance of the information for the tax process, decision-making should be carried out by the receiving organisation. Many requirements concerning the data format are legally specified, leaving little room for variation.

In Peppol, owing to its decentralised landscape, decision-making rights are more consensus-based and scattered. It is more challenging to enforce decisions, and each decision must be aligned with other parts of the PII. Decisions in Peppol or European governance structures cannot be made solely by national governmental organisations and have to be lobbied and influenced. This results in less control and increased uncertainty over the outcome. However, as respondent 2.2 noted, the Dutch government is very invested in this topic. The Netherlands is a frontrunner and can therefore influence many of the decision-making processes, as other organisations and countries are less involved.

The combination of empirical interviews and literature results in a framework for governance structures, as presented in Table 16.

Table 16. Information design variables

<i>Variable names</i>	<i>Design options</i>						
<i>Number of governance structures</i>	One	Multiple	None				
<i>Scope of governance</i>	Process	Format	Systems	Combination			
<i>Structure of the governance</i>	Hierarchical	Network	Heterarchical				
<i>Multiple levels</i>	Yes	No					
<i>Involved in Governance</i>	Public	Private	Both	None			
<i>Involved stakeholders</i>	Supplier	Receiver	Intermediate (p/p)	Regulator	Support	Control	Combination
<i>Decision-making rights in governance</i>	Public	Private	Both	None			
<i>decisions making process</i>	Power of perseverance by 1 (or a few)	Majority rule	Consensus based				
<i>How is the governance funded?</i>	Central funding involved organisations	Pay for participation	Central funding by a third party				

5.4 CHAPTER CONCLUSION

This chapter outlines a comprehensive set of design variables that collectively form a conceptual framework for PII. These variables are found in each of the data, technical, and governance components of a PII. For each component, we provided relevant variables and their design choices, providing a clear structure for understanding and comparing PII designs. Throughout the research, dependencies among variables were identified, such as how both decentralised and centralised infrastructures benefit from standards and can deliver services through intermediary organisations. It is valuable to explore these PII configurations in more depth and identify prevailing configurations in practice.

The analysis revealed interdependencies among the various design variables, emphasising that these elements should not be considered in isolation. For example, the type of data influences the technical choices available, which, in turn, affects the organisation of the PII governance. This highlights the importance of considering design variables not in isolation, but as part of a coherent configuration. The cross-case analysis of three PII cases illustrates that each case represents a unique arrangement of choices across the identified domains, shaped by contextual factors and stakeholder dynamics. These empirical findings suggest that while variability exists, patterns may emerge towards prevailing configurations in practice.

The next phase of this research focuses on identifying recurring patterns across a broader set of PII. By doing so, we aim to identify prevailing configurations that can inform both theory and practice in the design of PII.

Chapter 6

CONFIGURATIONS OF PII

In this chapter, we move towards phase three, *Configurations*. In the previous Chapter, we identified a set of relevant design variables that shape PII configurations. These include data, technical and governance design variables. For each design variable, several design choices could be made, resulting in a PII configuration. This highlights the importance of considering design variables not in isolation, but as part of a coherent configuration. We anticipate that prevailing configurations will emerge from the empirical analysis of PII cases. A prevailing configuration is a pattern of design choices that either occurs frequently across multiple cases, indicating a common or preferred approach, or deviates from other configurations in an analytically relevant way. In this chapter, we therefore answer the following sub-question: **What are the prevailing configurations of PII through which organisations exchange data with (semi-)public organisations?**

As established in Chapter 2.5.1, the hierarchical clustering method is the most suitable clustering approach for this research. A disadvantage of hierarchical clustering is that the scale and number of variables influence its results. Adding more variables can alter the clustering structure because each variable influences the distance calculation. When there are too many variables, noise can overpower the analysis, making it difficult to identify clear and meaningful clusters. Conversely, with too few variables, important patterns or dimensions might be missed, causing relevant clusters to remain hidden. Therefore, we decided to identify sub-configurations for each PII component separately. We start by identifying the sub-configurations within the data, governance, and technical components, as presented in Figure 6-1. We cluster the data based on the design choices made for these specific components. Data design choices identify sub-configurations within data components; technical design choices identify sub-configurations within technical components; and governance design choices determine sub-configurations within governance components. In addition to the governance sub-configuration, we focus on the individual governance structure, as some variables were not at the sub-configuration level, but at the individual governance structure level. The results of zooming in on a single governance structure are presented in Chapter 6.4

Using these sub-configurations, we compared the cases and their configurations to identify recurring patterns that lead to the prevailing configurations. We conclude that, based on the configurations found in Chapter 6.1 to Chapter 6.3, there are six overall prevailing configurations for the entire PII in practice, as presented in Chapter 6.5.

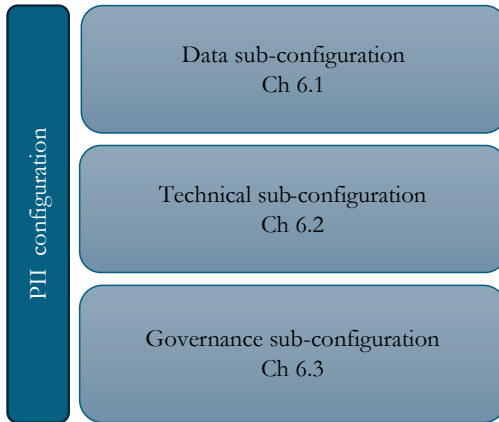


Figure 6-1. Configuration of a PII and its sub-configurations

In Chapter 6.6, we address the sub-research question: **What are the strengths and weaknesses of the prevailing configurations of public information infrastructures in terms of their quality?** The initial plan was to analyse the configuration of the entire quality framework. However, it became clear that a more in-depth study was needed to examine PII quality in detail. Therefore, we opted for a simpler approach to measuring configuration quality, while still aiming to make meaningful statements about quality. Therefore, we provide insights into the prevailing configurations’ strengths and weaknesses, considering perceived data quality, technical quality, and governance quality. We observe that each configuration has advantages and disadvantages. This highlights that one configuration is not necessarily superior to another, but rather depends on the requirements and context of the PII.

6.1 DATA SUB-CONFIGURATIONS

Through hierarchical clustering, we reveal the underlying patterns of data design choices within PII. A large part of this includes data formats. Data formats enable standardisation of data, ensuring interoperability and consistent data exchange between senders and receivers. As presented in Chapter 2.5, to uncover the underlying data design choice patterns within PII, we conduct a hierarchical cluster analysis. To establish the configuration of design choices of a PII, we used interviews, surveys, and document analysis.

This analysis aims to identify groups of patterns that share similarities. The clustering is formed by variables from the framework as presented in Chapter 5.1.2:

- The presence of structured or unstructured data exchange.
- The number of data formats in use.
- The reuse of existing data formats versus the development of specialised formats.
- Issues related to data format syntax, semantics, and technical implementation.
- The extent of collaboration in the design of data formats.
- The involvement of public and/or private sector actors in the design and management of data formats.

Figure 6-2 shows the dendrogram of the PII component data. The dendrogram shows how data-related choices are grouped by similarity. The figure illustrates six distinct configurations, each marked with different colours. The x-axis represents the cases, while the y-axis represents the distance between them. Cases that are different will have a distance of 1. The smaller the distance between cases, the more likely it is that these cases can be grouped into a cluster. For example, the cases VUM and SBR education are very closely related and have a similar configuration; therefore, they are levelled in the same cluster. The distance between these two and the case of WOCO was also small (lower than 0.2), so they fell into the same cluster. However, compared to the OOTS case, the difference is large. The case of OOTS resembles few other cases (distance > 0.8), forming its own cluster.

The final number of clusters was determined based on a combination of the dendrogram analysis, a visual inspection of the cluster matrix, and a qualitative assessment of the relevance of clusters.

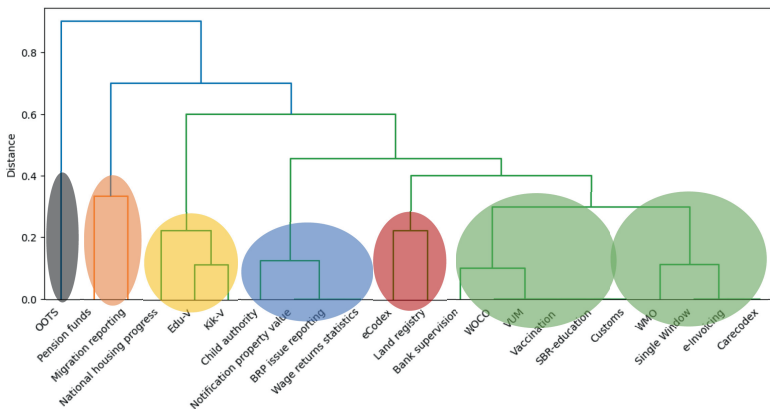


Figure 6-2. Dendrogram data configurations data models

In many clusters, a configuration was represented by multiple cases, but in two instances, only one or two cases represented the configuration. A possible explanation is that these clusters do not represent fully developed configurations, but rather precursors of other configurations, or they may be outliers that occur infrequently. Alternatively, they might reflect emerging types of configurations that have not yet been widely adopted. For this reason, we include all clusters in the analysis and discuss why only a few cases represent specific clusters. Additionally, it can be seen that the last cluster is larger than the others. Based on the dendrogram analysis and the qualitative assessment of the cluster relevance, we establish that these are the same type of sub-configuration; however, one cluster allowed one format, whereas the other allowed multiple formats. This difference will be explained in chapter 6.1.6 and 6.1.7. Table 17 presents an overview of the six data sub-configurations, the design choices characteristic of the sub-configuration, and the cases showing this type of sub-configuration. Each row represents a distinct sub-configuration, characterised by its own unique set of design choices. In some sub-configurations, we observe a single case, whereas in other sub-configurations, they are more widely found in practice. In the following subsections, we will examine each of these data sub-configurations, the characteristic design choices, and examples from the cluster.

During the analysis, we found some overlap among these configurations. For example, the eInvoicing case study features both a decentralised governance structure and a European focus, with a Dutch supporting committee. We therefore included it in both sub-configurations during the analysis.

Table 17. Overview of data sub-configurations, their characteristic design choices and the representing cases.

#	Data sub-configuration name	Design choices characteristic of the sub-configuration	Representative cases
1	Low structure and low in collaboration	Not structured	OOTS
2	Little structure and little collaboration	Structured by sender, Technical format, Hardly any collaboration	Pension funds, Migration
3	Semantic Focus	Structured, Focus on semantics, High focus on reuse, Public-private collaboration	KIK-V, EDU-V, and National Housing Progress
4	Public structure collaboration and reuse	Structured, Focus on collaboration, Focus on reuse, Mainly managed by public organisations	Wage returns statistics, BRP issue reporting. Child authority, Notification property value
5	Digitalisation of official documents	Structured Reuse by using existing (official) documents. Technical and semantical	ECodex, Land registry
6	A single fully structured collaboration	Structured, Specialised formats, Collaboration with stakeholders One format.	SBR-education, WOCO, Import, Vaccination, Bank supervision, VUM
	Multiple fully structured collaborations	Structured, Specialised formats, Collaboration with stakeholders, Multiple formats.	Carecodex, eInvoicing, SW, WMO,

6.1.1 DATA SUB-CONFIGURATION 1: LOW STRUCTURE AND LOW IN COLLABORATION

The first data sub-configuration represents a low-structured configuration, without clearly defined formats or technical standards. The OOTS case illustrates this sub-configuration. In this instance, there are yet no agreements on the data format. The data formats exchanged are bilateral agreements, and the focus here is on the technology and establishing the origin of the documentation, rather than the content. *"In OOTS, I do not care what you exchange, as long as you know it comes from an actual data source."* (R3.28). At present, the predominant form of message exchange is through a PDF structure. This case is in the stage of converting paper-based processes into digital exchange. The first stage involves mapping the data. The subsequent investigation will determine which data is both available and requested. *"That is identification. So, we consider whether, for this requirement with your procedure and the target country, you have a one-to-one match between your requirement*

and existing evidence types, or if you need to combine them from those two or even multiple sources.” (R3.28). There are several initiatives aimed at transitioning to XML for documentation. *“You can see that countries that are close to each other, like Germany and the Netherlands, have quite similar data in use”* (R3.28). This is then a situational XML project. We anticipate that, over time, more initiatives of this nature will advance towards structured message exchange.

6.1.2 DATA SUB-CONFIGURATION 2: LITTLE STRUCTURE AND LITTLE COLLABORATION

Data sub-configuration two is characterised by a unilaterally imposed data format. Both the Pension Funds and Migration data cases involve dumps from existing processes, thus reusing existing data. The data is derived by subtracting from existing datasets, shared with a third party without modification. The data recipients are only minimally involved in the data formation process.

The data format is characterised by minimal structuring and is often a CSV or Excel file. *“Well, most datasets are classically delivered via CSV files. The government in this form, with these chain partners, is not yet very modernised in data exchange.”* (R3.19.) This raises the question of whether the process would benefit from structuring the data or if this creates additional workload with limited benefits. Nevertheless, there is close coordination between the parties regarding the interpretation of the datasets. *“We always ask the delivering party(...) whether we have interpreted your information correctly”* (R3.19). After all, data without context cannot become meaningful information. Even when changes are made to the dataset, this is coordinated with the receiving party. *“The format was once established in the beginning, and if a change to the format happens there, it does get done in consultation.”* (R3.1) It is characteristic of this data sub-configuration that there is an absence of effective coordination concerning the structure of the data. It appears that effective communication and collaboration between the relevant parties in subsequent data-related processes are of greater importance.

6.1.3 DATA SUB-CONFIGURATION 3: SEMANTIC FOCUS

The third data sub-configuration is characterised by an emphasis on the interpretation of elements, with a particular focus on semantics. Examples in this sub-configuration include the cases of KIK-V, EDU-V, and the National Housing Progress. While reuse is possible, it is not a prerequisite. *“So, what a healthcare provider sets up is a data set, based on an ontology. We set up a data dictionary. (...) There are all kinds of data defined that are already recorded at the healthcare provider with certain meanings, so what we expect that data to mean.”* (R3.15) It is argued that collaboration between different parties is highly important. *“There is a governance structure (...) for the policy, so the ministry and municipality. All kinds of parties are included, with a basic set as a result. So recently they had the basic set 2.0 again, (...) In cooperation with those 35 municipalities, of course, we have always coordinated about what information they would like to see in it.”* (R3.16). This coordination is considered important to ensure data quality. *“So, they also benefit from it that the data from municipalities is as high quality as possible.”* (R3.16) Engaging with each other on the content of the data can initiate discussion about its potential applications and encourage critical thinking while formulating questions. *“So, we are now trying to make the questions [answered by the dataset] as clear as possible and therefore comparable. That also solves a political issue, because now it just depends on which number you get from where*

and how you want to get it right. Anything is possible, because there is total ambiguity, [about the data] (...) So, in the discussion, we try to not only write these kinds of values, but also to be able to provide this kind of context. But also, especially clarify the question in such a way that it can really only be interpreted in one way from the answer. Or ask multiple questions if there are multiple questions in it, because sometimes there are as many as five questions hidden in one question." (R3.15). In addition, EDU-V has established a working group that is engaged in the active development of definitions used within the PII. *"We start with a conceptual model for learning and teaching resources and activity data. Based on this conceptual model, the data model was worked out in detail. Finally, definitions were worked out for each domain."* (R3.8). We therefore observe a heightened emphasis in this configuration on the content of the data and collaboration to foster mutual understanding, rather than on the technical aspects of data.

However, it is important to recognise that these cases are in the early stages of the maturity model. This may indicate that this sub-configuration is a preliminary phase before the formation of data sub-configuration six. The initial phase involves identifying exchangeable elements and establishing a suitable format to ensure their integration into the infrastructure. Nonetheless, EDU-V and KIK-v differ in how they exchange data and could represent a new kind of infrastructure that shifts the focus from syntax to semantics, because data is retrieved rather than transmitted.

6.1.4 DATA SUB-CONFIGURATION 4: PUBLIC STRUCTURE COLLABORATION AND REUSE

In this fourth data sub-configuration, the emphasis is placed on collaboration and the reuse of data. It is important to note that these configurations involve exclusively data exchange between public organisations. The data sent through the PII originates from existing processes and is reused to inform relevant parties. For example, in the BRP issue reporting case, errors can be reported in the original data by sending a subset of the BRP. In the context of payroll declaration, the subject can be seen as a subset of tax authority data. In the Notification property value case, it involves a subset of changes to property values, and in the Child authority case, it entails sharing a subset of case documents related to a family or individual.

In all these cases, there is only one data format. The case of BRP issue reporting utilises a data 'dictionary' called Catalogi. *"To ensure a reporting organisation reports the right thing, we have a data format. This tells us the object types, attribute types, and other details necessary. (...) If you report back to the BRP, you report on personal data. (...) for this you require the Catalogi"* (R3.2). In the Child Authority case, respondent 3.11 noted that it starts with an information model, and based on the collective understanding of the information, data exchange profiles are made. This is why it is a distinction of configuration three, where the focus is more on understanding. *"We specify the notification, the command API and the perusal API. And then we ask all parties, who have notification software or case package software, for example, to be able to receive that notification, send out the notification, provide the perusal and so on."* (R3.11). These cases not only prioritise the content but particularly emphasise the format for exchange.

The content of the information model results from a collaborative effort between the sender and the receiver. Conversely, data exchange profiles are often developed through collaboration with intermediary organisations. Compared to the first two data sub-configurations, these cases involve greater collaboration.

Part of this stems from the legal obligations of the relevant organisations and the involvement of intermediary organisations in the process.

Creating an information model can be challenging due to the growing number of actors in an information infrastructure. In the Child Authority case, the respondent (R3.11) observed that many data items are shared among more than 4000 organisations. This demands extensive collaboration to agree on a single information model, which also requires changes in each organisation's processes. As the respondent noted, *"For example, the police have their own language in the police domain. However, if they want to collaborate in the youth and family domain, they have to speak the shared language. (...) Systematically, it gets complicated, with 4,000 parties- such as 400 municipalities, 600 neighbourhood team-like groups, and 4,000 youth care institutions- needing to adopt that common language in their systems. So, whether that will work, I do not know"* (R3.11). In some cases, it's therefore impossible to establish a single format and dictionary where each part defines concepts in the same way.

In this configuration, the development and management of the data format are responsibilities shared by multiple parties, each influencing the content of the message. The involved parties in the data exchange have the capacity to specify the type of data to be shared. This raises concerns about data responsibility. As respondent 3.17 of the wage return statistics case noted: *"Originally, it is a financial data stream from the Tax Administration, it is the pay slips. But there are just parties that would very much like to connect to the payroll declaration chain. As a result, additional variables were needed. Think of the pension funds. (...) Then there is the question of who owns that data. The Tax Administration says they do not want the responsibility for additional data, as they cannot check the quality of those variables."* Data ownership is especially important to establish and clarify among all participants in the information infrastructure.

6.1.5 DATA SUB-CONFIGURATION 5: DIGITALISATION OF OFFICIAL DOCUMENTATION

Data sub-configuration five can be viewed as an intermediate cluster. In this specific sub-configuration, official documents are being digitised to some degree. Both cases in this sub-configuration are in the process of (partially) digitising their processes. There is consultation between the different parties, and there is (to some extent) a structured message format. However, this is not yet utilised to its full potential, and collaboration remains in its early stages. There is no re-use, as existing legal texts or notarial acts are used as message exchanges, which are purely digitalised. *"There is a common denominator in all these exchanges. That is, each exchange must be based on an existing legal instrument or regulation. So, it should never be a question of whether another country can actually receive the information you send. The existing regulation says they do."* (R3.7).

Consequently, there is limited discussion about the content of messages. *"The simplest example is still that the first word of such a style sheet is 'Present' or 'On' or 'Today.' An action should always start with one of these three words."* (R3.13). According to the respondent, the Land Registry has undertaken the digitisation of a small proportion of relevant documentation, but not all of it can be digitised. It is important to note that a significant amount of information remains to be completed by the notary. In the case of eCodex, it is clear that harmonisation of messages is occurring at the European level. In contrast, within the Land Registry context, this process is conducted at the national level. This is related to the actors involved in data exchange.

6.1.6 DATA SUB-CONFIGURATION 6A: COMPLETELY STRUCTURED SINGLE FORMAT

Data sub-configurations 6A and 6B both represent developed data formats with official governance structures. The key difference between them is the number of data formats they can deliver. In sub-configuration 6A, only a single data format is permitted. Additionally, all message structures are based on a higher-level framework with further details at the message type level. Both start with the semantics: *“If you are going to exchange data, you need information. You need to know what information you need, and then you also need to have the same interpretation of that information. This means you have to make agreements about that beforehand. So, we worked out with the parties what information you need. To be able to make a match.”*(R3.23). We observe that, in many cases, semantic models or taxonomies are utilised. *“I know that there are two main types of taxonomies that are widely used at the European level. So, on the one hand, you have EOPA. That is the institution for pension funds and insurers. They actually create a taxonomy.”* (R3.4). SBR-education and WOCO also employ a taxonomy; both use the SBR taxonomy, which can be supplemented with organisation-specific elements. *“But XBRL is linked to the Dutch SBR standard, and that is just one taxonomy. Internationally, we have adopted that. Just like the next phase that we are going to do, IXBRL. We did not invent that. That already exists. Internationally. It is just that it is now being introduced and made mandatory for large institutions in the Netherlands. They are working on it. And we are following that in the world of education. When they are ready with that, it will be part of the KV/K that does that. That is really just the basis. And then we say, when that is done, we will take that. And then we do our own little things.”* (R3.6) Here, we observe the reuse of existing formats and techniques to form data, which are the core of the sub-configurations, but the key is that they reach a common understanding of the data format.

This also makes it easier to compare data sets across data streams. *“They decided once to join SBR. Standard Business Reporting. There are nice taxonomies in there. It once started at the tax department for accountability, but they copied those for housing associations. If you can reuse the basics, you have to make good agreements of meaning in other taxonomies?”* (R3.24). With the bank supervision case and the import declaration process, many data formats have been harmonised at the European level, and at the national level, it is possible to request additional data. Thus, each country has its own implementation of the taxonomy, leaving room for extra requirements in each nation. As reported by respondent 3.9 in the case of import declaration, there are currently 26 Member States, resulting in twenty-six different connections and message types for import. However, it is clear that this process does not stand still but is constantly evolving to improve and optimise: *“We are now working on the successor. It is called Customs reform. It will be watertight within the next 10 years. Then it really must be the same on every level. Because we want that within the Netherlands, the declaration of importing goods is the same as in France. (...). This gap is as good as being closed. However, it is not exactly the same at the moment.”* (R3.9). This demonstrates the importance of continuous development and harmonisation of data (formats) across all levels of the PII.

The importance of collaborations on semantics and data content is emphasised here. Collaboration is essential here, and prescribing alone is insufficient. *“Suppose you have a digital highway, and you do not invest in good collaboration that ensures that you also agree on what you are going to provide. Because that always takes time and involves work. If you don't, you are just going to get the same crap you are getting now, only faster. You do not want that. So, some form of cooperation is really going to be necessary.”* (R3.5) Therefore, it is evident that this is a collaboration between

public and private parties in designing the message format, and many cases within this configuration are very inclusive. An example is the case of SBR-education and WOCO. Both the SBR-education and the notification property value cases are part of the same governance structure, namely SBR, which stands for Standard Business Reporting. This is a large and inclusive governance structure that involves both public and, where applicable, private representation. Within this governance structure, multiple parties have come together to not only harmonise the content and meaning of messages but also standardise the method of delivery.

However, in other contexts, we observe that standardisation does not end with a single message type. Regarding Vaccination data, the creation of the standard involves collaboration among various organisations. In this case, standardisation groups such as NEN and a dedicated standards organisation are involved to achieve harmonisation. *“This is where information standards are made, and NITCIZ still does that. (...) There are even NEN standards for that, so every party involved can also participate in governance. That is the NEN 7522. A NEN standard that addresses the governance of information standards. (...) This includes the data exchange of personal files, the data guideline of the National Vaccination Programme, and the National heel prick screening.”* (R3.2). The message format must be developed through a collaborative effort between the public and private sectors. Additionally, it is important not only to establish the semantic model during these meetings but also to consider additional syntax bindings. This practice ensures consistency, harmonisation, and interoperability. The shared standard for data exchange is XML. In the context of Bank supervision, WOCO, and SBR education cases, formats such as XBRL are used. This is a more complex variant than XML, ensuring that data can be quality-checked through additional rules in the format itself.

Using a more complex standard can have both positive and negative effects. To illustrate this, one respondent highlighted the risks associated with adding complexity to the data format. *“It is only a few small parties that know XBRL. Moreover, who can offer XBRL tooling? That market is sparse. So, if you want to get rid of that party, that is difficult. Then you end up with the same party again anyway. If there are thirty software vendors, there are not that many who are also XBRL experts in the Netherlands. Those can be counted on one hand. That is a few parties and a few people. Moreover, everybody knows everybody. So that is not dozens of companies or thousands of people who know the trick.”* (R3.24). However, another respondent approved of XBRL and the reuse of existing taxonomy structures. The respondent 3.6, in the SBR-education case, noted that this approach involves applying established technological solutions. This way, existing software vendors, who have already implemented these solutions in different ways, can adapt them for a new application area. This means you can use current technologies by modifying them, rather than starting from scratch.

6.1.7 DATA SUB-CONFIGURATION 6B: COMPLETELY STRUCTURED, MULTIPLE FORMATS

Data sub-configuration 6B is a variant of sub-configuration six that allows multiple data formats. Different syntax bindings (and technical formats) can be chosen while keeping the same semantics. The data shows that technical legacy and diverse infrastructures are often the main reasons for supporting multiple data formats.

The eInvoicing case involved both market-established formats and government-created formats. Within the governmental systems, data formats are converted from market-established to government-established formats. The format used depends on how the data is provided; using the portal or a direct connection to the government systems typically results in the use of government formats, while employing a private intermediary or the Peppol network generally leads to the use of market formats, as stated by respondent 3.3. The Case of WMO also has several data formats *“And besides managing the infrastructure, we are also responsible for various standards. We have several standard administration protocols, contract standards, and various handbooks in the field of procurement in the Social Domain (...). We create i-standards. The interest of municipalities in these standards increases as well”* (R3.10). This highlights the growing standardisation of data formats in the public domain and the recognition of the added value of standardisation in the PII.

Interestingly, having multiple data formats is not always viewed as an advantage. *“One could ask what is complex about a financial administrative process. And why is it different in Westervolde municipality than in Vlissingen municipality? Because in the end, it comes down to the same thing. The care provider simply has to be paid for the care that has been delivered. And there are variants, but you also have to implement them properly (and consistently) and not mix up variants.”* (R3.10). The variety of formats often involves legacy systems. For example, in the case of SW, existing systems often used Edifact, whereas newer systems used XML. Yet, the declaration to Customs is just a small part of the import process, and changing all data exchange to XML would require updating all earlier systems in the chain as well. This process is often costly and time-consuming. It is not always aligned with the objectives of market participants. To address this issue, governments often authorise the implementation of different data formats.

This also creates a business case: a new area of expertise develops as companies specialising in data formatting and transformation emerge. The primary function of these companies is to facilitate data sharing between organisations and to provide interoperability within the network.

6.2 TECHNICAL SUB-CONFIGURATIONS

In this section of Chapter Six, we perform a second analysis using hierarchical cluster analysis to understand the underlying structure of technical components, including systems, services, and standards, within PII.

In the context of PII, a service refers to a functionality offered during the data exchange process, a system is a collection of connected hardware and software, and standards are agreed-upon specifications that ensure compatibility and interoperability between different systems and services. There are multiple levels at which design decisions occur. This chapter focuses on technical design choices at the level of the entire PII, rather than a single system, service, or standard.

This analysis aims to identify patterns and groupings based on shared characteristics across various infrastructural setups. The clustering is formed by variables from the framework as presented in Chapter 5.2.2 and includes:

- including the type of infrastructural relationships (e.g., 1-N, N-1, N-N, 1-1),
- information flow mechanisms (push or pull),

- and principles such as ‘send once, use many’.
- structure of the infrastructure (centralised, decentralised, distributed),
- the presence of single or multiple systems,
- involvement of public or private sector intermediaries,
- the extent of collaboration or reuse,
- and the use of (open) standards.

The variable of the included services in the process of the data exchange was not included in the analysis of the technical sub-configuration. There were too many services to include as variables, causing problems in the hierarchical clustering method. Furthermore, it was unclear which variables appeared in the different PII processes, as some systems in the PII cases did not offer a clear overview of the services provided.

Figure 6-3 shows the dendrogram of the PII technical component. The dendrogram shows how the technical-related variables are grouped based on their similarity. The x-axis represents the cases, while the y-axis represents the distance between them. Cases that are entirely different will have a distance of 1. The smaller the distance between cases, the more likely it is that these cases form a cluster.

For example, the cases Single Window and WOCO are closely related and share the same configuration; therefore, they are grouped in the same cluster. The distance between these two and the case of BRP issue reporting was also small (lower than 0.5), so they fell into the same cluster. However, compared to, for example, the Pension fund case, the difference is large. The case of Pension funds resembles little other cases (distance > 0.8), forming its own cluster. To obtain the final configurations, we use both the dendrogram and a qualitative data analysis. This combination ensures that the cluster not only reflects statistical similarity but is also grounded in the contextual nuances and insights derived from the qualitative interviews.

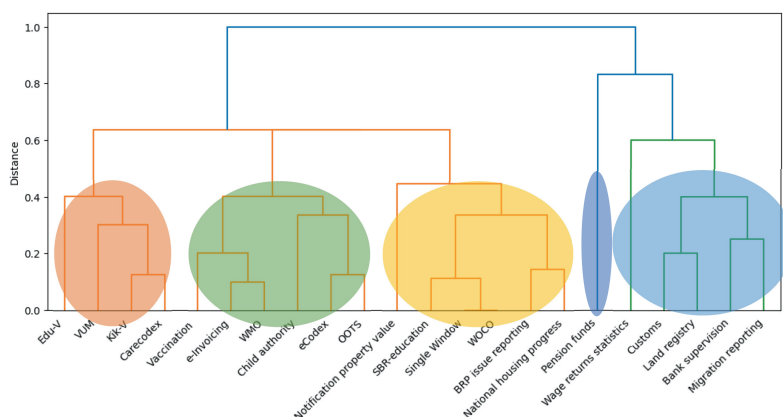


Figure 6-3. Dendrogram of configuration in technical variables

Based on the dendrogram analysis, we identify the different clusters with a distance value below 0.6. The pension funds show to be an individual sub-configuration. However, during the qualitative assessment of the cluster relevance, we established that these are the same type of sub-configuration. The difference is that the pension fund involved a pull mechanism, whereas the other involves a push mechanism. This difference will be explained in Chapter 6.2.1. We therefore establish that there are four distinct technical sub-configurations, of which one sub-configuration has two variants.

Table 18 presents an overview of the six technical sub-configurations, the design choices characteristic of the configuration, and the cases showing this type of configuration. Each row represents a different technical sub-configuration, with its own characteristic set of design choices. In some sub-configurations, we observe a single case, whereas in other sub-configurations, cases are more widely found in practice. In the following subsections, we will examine each of these technical sub-configurations, the characteristic design choices, and case examples in the sub-configuration.

Table 18. Overview of technical sub-configurations, their characteristic design choices and the representing cases.

#	Technical sub-configuration name	Design choices characteristic of the configuration	Representative cases
1	Central infrastructure with minimal collaboration	Push mechanism, N-1, Central system, Limited number of intermediaries, Low level of collaboration,	Migration, Import, Bank supervision, Wage returns statistics, Land registry
		Pull mechanism, 1-N, No intermediaries, Central system, Low level of collaboration	Pension funds
2	One central infrastructure	Push mechanism, Three variants: N-1, 1-N, and N-N; High level of collaboration, Multiple systems with one central node, Presence of intermediaries,	WOCO, SBR-education, National housing progress, BRP issue reporting, SW, Notification property value
3	Hybrid infrastructure	Hybrid / (semi-)decentralised, N-N, Primarily public intermediaries, High level of collaboration around shared parts of the PII,	ECodex, OOTS, eInvoicing, WMO, Vaccination, Child authority
4	Distributed infrastructure	Distributed infrastructure, sometimes with a temporary central component. N-N High collaboration, Both public and private intermediaries.	Kik-V, Edu-v, CareCodex, VUM

6.2.1 TECHNICAL SUB-CONFIGURATION 1: CENTRAL INFRASTRUCTURE WITH MINIMAL COLLABORATION

The first technical sub-configuration is a central infrastructure with minimal collaboration. It is characterised by a centrally managed infrastructure and a limited number of intermediaries, if any. This mainly involves infrastructures where multiple senders exchange messages with a single receiver. A push mechanism is used to exchange data, with little to no cooperation regarding the systems in the infrastructure. These infrastructures are enforced unilaterally. In the following paragraphs, we examine the defining characteristics of this configuration in detail.

In this configuration, there is typically an N-to-1 relationship; here, the receiving party dictates how the message exchange takes place to the supplying parties. In cases such as Migration, Bank supervision, Import, and Land Registry, multiple supplying parties deliver directly to the receiving party. As the recipient, they can specify the technical requirements that the supplying parties (or their representatives) must follow for delivering data. However, in the case of the Wage Declaration, it is a one-to-one relationship. Here, the supplying party determines how the exchange takes place. As the recipient, they rely on the data and the data suppliers. Therefore, they need to be more flexible (R3.17).

Consequently, existing systems are rarely reused within this sub-configuration. *"We are in charge ourselves. We also develop a lot of software ourselves. So those things we just mentioned, like KIK, we developed ourselves. Heading, that was really a source system. We developed that ourselves, too. BK Notification was developed in-house. We developed the submission process from scratch. Of course, we do use building blocks from outside sources. So, we do use existing software components. If you didn't do that, then you would be very strange."* (R3.14). Systems are often primarily developed specifically for individual organisations.

Based on the data, we can deduce that these sub-configurations utilise existing technical standards and components to develop the system. Examples include tooling for qualified signatures, the login method, e-recognition, and event sourcing. According to the Land Registry respondent, 'reuse before buy before make' also applies to the development of their own systems. However, it should be recognised that many of these processes feature unique components that are not available on the market. The respondent 3.13 from the Land Registry case stated: *"We are the only ones in the Netherlands who do that. So why provide a standard package on the market? There is no market for it. We will just build it ourselves."*

One notable feature of this configuration is the use of portals. Suppliers can upload data through a portal for data exchange purposes. In the case of import declarations, data can be entered once an importer or declarant has been granted access. During the migration, access is provided to a central data platform, and at DNB's infrastructure, service reports can be submitted in the MijndNB environment. This is offered by the receiving party itself. The portals used in bank supervision and import declarations also display status information regarding message delivery.

A variant of sub-configuration one concerns pension funds. In this case, a portal is also employed. *"We have to log in to the UWV portal. That is where the data is available to download. We can download it from there. Then we have to prepare it internally in our processing system. Logging in is done with e-recognition at level 4. That is the highest security*

level within e-recognition” (R3.1). However, here the receiving parties retrieve the data from the supplying party, establishing a one-to-many relationship. In this case, the enforcing party is the delivering party. *“We check whether the BSNs they send are correct. We also check if it is technically processable”* (R3.1). The pension funds rely on data from the delivering party that complies with the UWV, and they perform a data quality check.

All services are offered by the receiving party, who also supplies the system to be delivered directly. Therefore, the responsibility for these services lies with the receiving party. Data quality control is an example of this. *“The applications that check the declaration. The input declaration is checked for syntax. Is it all filled in? Is it filled in correctly? Then we validate things.”* (R3.9). The end-user also validates the content when applying the data. *“Of course, we also compare that information. What is published? What reports are there? Does that say the same? So, it is mostly verification-based. Every chain partner of mine produces their own reports. If a party says they have ten people in the shelter and we receive another file stating eight people are in the same shelter, we have a conversation about it.”* (R3.19). In this configuration, this is carried out by the receiver, which can be achieved through technical validations.

In this technical sub-configuration, minimum requirements are imposed on the submitting parties to facilitate their submission process. The respondent 3.19 (Migration) mentioned that they are content that resources are allocated to data exchange at all. The migration chain lacks a binding framework and therefore relies on the supplying parties and whether they prioritise it.

In the bank supervision case, an in-house service is designed to convert the data to XBRL format. *“We receive that directly through a service that we have ourselves. Then it is actually validated. This is external software that we hire. It parses the entire XBRL file into the correct format. Then we store it ourselves. DNB built all that itself.”* (R3.4).

As shown, the responsibilities in the process are clear. The supplying party is accountable for the data content. The party receiving and managing the system is responsible for the technical systems and data processing.

In some instances, a party employs private intermediaries to facilitate data exchange. In this configuration, it can be noticed that they only offer technical support for the exchange, not additional services such as data quality control, data format transformation, or signature services. *“When we first started working with Digital Highway, it was useful that RINIS possessed centralised knowledge for the safe and fast transport of data, even for very large files. That was not nearly as obvious then as it is now. So far, these exchanges via RINIS have actually operated without any issues. There have been hardly any malfunctions in the past twenty years. Therefore, there is no immediate need to change the approach. Certainly not while ongoing processes are in place. Once you have established your processes on both sides, you are not going to change them just for the sake of it.”* (R3.17). According to the respondent, the CBS's objective within this configuration is to keep the data exchange chains as short as possible. The respondent from the Migration case also noted that connections and technical exchanges are supported by the private sector. However, the data exchange takes place between organisations within the public sector.

In cases such as Import Declaration, bank supervision cases, and Land Registry, private parties are involved, representing the sender. *“It may also be that they ask an intermediary, a Customs broker, (...) I would like you to make the import declaration, because that is complicated, and I don't feel like it.” Just as you used to have your tax form filled in by a tax consultant, you can do the same. (...) That is just a service”* (R3.9). In this case of import declaration,

there are brokers who make import declarations on behalf of traders. The importance of private parties in this sub-configuration is therefore high.

In the bank supervision case, the auditors are responsible for providing the reports. This requires accountants to issue an audit opinion to validate the reports' content. Finally, within the Land Registry, notaries are responsible for delivering the deeds. While some cooperation does take place, many conditions are still enforced by the receiving parties.

This technical sub-configuration seems generally stable within this type of infrastructure. For some cases, it signifies the start of digitisation, while others have already been operating with this configuration for some time.

In the case of the land registry, the digitisation process is still in its early stages. *“Software suppliers also have to invest. They have to finance all these modifications to their system themselves. So yes, that also influences what you include and exclude.”* (R3.13). The respondent added, *“Because it is so free, it is an option, not an obligation. You can see that it is not really embraced either.”* (R3.13). The key question is whether there is a sufficient basis to develop the PII further. In the payroll declarations case, respondent R3.18 noted: *“If the (data exchange) chain is running well, then the need to meet each other is also reduced. However, to remain robust, you do need to consider how we proceed. But that depends on the priorities within the organisation itself. As long as that is recognised and addressed, the other parties will too. If that is made clear, then we hope it all continues to go well. Of course, there is no guarantee for the future.”* (R3.18). This case has been configured this way for quite some time; there is currently little need for further development. However, without development, technical debt might arise, and new developments may be harder to implement. Any further development is heavily influenced by the level of (political) attention and urgency that other stakeholders within the PII assign to it. *“It is difficult to make significant progress when they say, ‘We have other priorities’. You would actually want there to be sufficient continuous capacity to move the chain forward.”* (R3.17). However, some cases with this configuration do develop, for example, during the delivery process. The respondent from Migration stated that they are continuing to develop the interfaces. *“We are now working on several innovations. We are doing that first with some smaller external parties, moving towards APIs and real-time links. We are still generally not retrieving individual actions with this, but rather bulk datasets in JSON, for example.”* (R3.19).

6.2.2 TECHNICAL SUB-CONFIGURATION 2: ONE CENTRAL INFRASTRUCTURE

The second technical sub-configuration is characterised by a central infrastructure provided by intermediaries. It involves a push mechanism for exchanging data, with a high level of cooperation regarding systems and standards. In this sub-configuration, infrastructures are established collaboratively within a governance structure, with representation from various parties being standard practice.

At the front end, the landscape is often supported by private software vendors or market players. These vendors ensure that data sharing organisations are supported in their business administration and the data exchange process. *“They often have a software package supplier where the whole planning, accounting is in which they make an interception to deliver S2S.”* (R3.6). It can be observed that within central facilities, there is an emphasis on adhering to public standards. These standards are developed collaboratively between public parties and the

public intermediary. However, there is a trend to include market parties as well, and the focus is shifting from public standards to more market-based standards.

The infrastructure primarily consists of a central system that receives data. In all cases, this involves collaboration between public bodies. Of course, different software systems can support sending or receiving parties in their processes at the front or back end, but this is usually managed separately. As respondent R3.24 noticed: *“We just have a generic solution for generic digital infrastructure. It is a public-private infrastructure that we are simply going to use. Then we will use generic building blocks. That is actually no longer an issue. You are not going to build a completely new one from scratch. You just grab what is available.”* Therefore, a key feature of this technical sub-configuration is the use of a shared system by multiple parties in similar processes.

The central system utilises common services, including a portal service, a conversion service, and access and authorisation services. Many of these services can be organised modularly within the system. This allows different parties to follow their own processes while sharing the same system. There was a period when designing central systems was quite popular. However, one respondent noted: *“I also come from a time when centralisation of those kinds of services and development was very much in the hands of the government (...) Maybe we should grow into a new model. Only that function of development that is now a bit divided among those different partners.”* (R3.5). This may indicate that this view is changing. A respondent in the notification property value case noted: *“You cannot think everything out in theory. Of course, you have to do a lot; you cannot just run blindly in one direction. However, you should not overthink everything in advance either. That is what you get now, though. All possible options are explored, but of course, nothing is right. By the time you have made up your mind, the tech world has already innovated and changed. You are really shooting at a moving target.”* (R3.25). Here, we observe that the landscape of central systems is evolving, but it is difficult to identify a new alternative

Not all PII are as mature. During the design or further development phases, we observe collaborations emerging to achieve harmonisation and standardisation. For example, in the National Housing Progress case, some parties require tooling for data delivery, while others already have systems with tooling. The respondent noted the need for uniformity in the tooling. Therefore, some parties are working together to build a shared system, while a broader group is collaborating to develop standards that enable the tooling to be applied more universally.

However, as a consequence of shared systems, coordination is needed to align stakeholders’ needs for the system. For instance, the Single Window respondent stated: *“If parties want something, for example, technical functional data changes, they raise it at one of the consultations. Then we have a discussion with the parties about what this means. What needs to be changed? Can that be done? Do they want that? Is it allowed? Do the parties want that? What about the costs?”* (R3.22). This means that additional consultations are required across the system for part of the chain or network.

In most cases, this sub-configuration is provided by a third (semi-)public organisation. *“Some central systems are used like Digipoort. That is a matter of public law. You do have to pay for that. You could do things differently, but then the market would also have to find an alternative itself. There are not that many alternatives”* (R3.24). In many cases within this sub-configuration, we observed that Logius functions as an intermediary. Logius is a government organisation that provides digital facilities, standards, and systems to support government

digital services. *“Logius provides systems and services such as portals. They co-developed the standard and provide Digipoort. Logius is also on the governance steering committee.”* (R3.24). It is an agency of the Ministry of the Interior and Kingdom Relations (BZK). In several cases, a public-oriented party is responsible for providing central systems, such as organisations like JUSTID or RINIS. In the case of the Single Window, the receiving party also acts as an intermediary.

Regarding the use of private intermediaries, respondents noted that this can be challenging. For example, the Single Window respondent stated, *“The price can start to get very high if you only have one party supplying it.”* (R3.22). Often, there is no viable business case for the parties involved. *“You can't earn that much from it either, and of course that's also dependent on the market. They have to invest. (..) It is a niche market, very small. (..)The market is not that big”* (R3.5). The respondent continued, *“As a government, we have to be very careful not to distort the market.”*(R3.5). Without a viable business case, private sector involvement will remain absent.

We can identify three variants within the central infrastructure. In variant one of sub-configuration two, multiple parties submit to a receiving party. This applies to National Housing Progress and BRP issue reporting. In this case, multiple parties deliver at the front end, and systems are often developed to integrate the delivery process into business workflows. Intermediaries at the front end of the chain assist with the transformation of datasets and provide portals or system-to-system links for data delivery.

The second unique variant is that of the Notification property value case. This involves a push from 1-N. *“This is a mutation message that should be distributed to all subscribers. Then you use Digilevering. This allows messages to be spread across lots of different recipients at lightning speed and is very stable and reliable”* (R3.25). In this case, it also involves sending a large number of messages in a short period of time. *“I cannot see how parties could process fifty million messages at once. There must be something like throttling or load balancing. That is included in the OCBK. The aim is to ensure that no messages are lost. So, you need a kind of queuing mechanism”* (R3.25). The Digileveren service provided by Logius is used for this purpose. The respondent noted that more infrastructures use the Digileveren service.

In variant three of this configuration, there is an N-to-N relationship. This means that multiple parties can deliver and receive messages. For example, in both the WOCO and SW cases, the central facility has a copying function. *“You send the message to the Digipoort, which copies it. It then sends the message to each party. The 350 parties who send in the DVI and DPI are copied and distributed to the three or four receiving parties”* (R3.24). In the case of WOCO, all receiving parties receive the messages sent in. In the case of SW, it depends on the type of message sent. *“It distributes that based on the matrix of interest, as we call it, to the different authorities behind the Single Window, and we then deliver that data to A, B and C”* (R3.22). Not every element in the data exchange is necessary for each receiving organisation. *“Depending on your role with respect to the data, there has to (...) be a legal basis for a government party to receive it. If there is no legal basis, that data should not be shared”* (R3.22). Organisations need to have a clear purpose for using the data element to be able to receive it.

6.2.3 TECHNICAL SUB-CONFIGURATION 3: HYBRID INFRASTRUCTURE

This third technical sub-configuration is characterised by the hybrid nature of the infrastructure. There are multiple systems and nodes within the infrastructure, some of which are centralised and some decentralised. A fully decentralised infrastructure is not present in any of the cases. From a government perspective, there are more central nodes in place, but the infrastructure still has a decentralised character.

Cooperation within these chains is crucial. Collaboration has established central nodes within the network. It is also noteworthy that these types of infrastructure are involved in message exchanges on both sides. In the previous sub-configurations, information flows primarily from the sender to the receiver. The receiver does not play a sending role through the infrastructure. In these cases, however, the receiving parties can also act as senders. The infrastructure is used more extensively for message exchange with all parties.

This third technical sub-configuration involves a complex hybrid infrastructure. This infrastructure can include centralised, decentralised, and distributed components. Decentralised components occur within the infrastructure in a few cases. For example, in the case of OOTS, a national facility has been developed to connect to the decentralised OOTS infrastructure. *“The idea is that everything goes through the access point at JustId to the other access points in other countries”* (R3.21). Additionally, within eCodex, there is a national facility that parties can connect to for exchanging intra-community messages. *“We have said that we have one access point. The complexity within our own country is our responsibility. Other countries only have to send messages through that access point. The ECodex access point is at the Judicial Information Service”* (R3.7). In both cases, this is offered by an intermediary. Ultimately, they are all connections between different countries and organisations within the system. Within eCodex, countries can choose to install a central hub or opt for more access points. When countries choose not to establish a nationwide intermediary facility, the infrastructure becomes more distributed.

In the case of eInvoicing, a central system is in place that intermediates to send invoices to the central government. There are multiple ways to provide the data to this central system. *“We have a supplier portal within the central government, particularly meant for the slightly smaller suppliers. Suppliers can manually create their invoices in the portal and click Send. And then that goes from the supplier portal through the Digipoort to the central government directly or through the central government's intermediary”* (R3.3). There are also market parties that can exchange data with the central facility via a link and technical standards. Additionally, there is a hub on the central facility that connects to a distributed network called Peppol.

Another form of hybrid infrastructure involves the WMO and Vaccination. In these cases, each domain has a central hub (node) that facilitates data exchange. These domain-specific hubs are interconnected, creating a decentralised infrastructure. Additionally, connecting parties use software packages that must link to the common nodes. This dramatically increases the number of stakeholders within the infrastructure.

This complexity appears to have arisen from collaboration among stakeholders within the domain, such as public entities collectively developing infrastructure. *“With the decentralisations of the WMO and Youth Act to the municipality, a hub was set up for the municipalities. So, all municipalities are connected to it. Moreover, on the care side, we have connected all providers to VEZOZO who provide care in the WMO and Youth Act”* (R3.10).

On the other hand, we observe that the existing infrastructure is not being completely replaced; instead, specific parts of the PII are being further developed. For example, connection to Peppol's distributed network is a recent addition in the eInvoicing case study. Similarly, eCodex and OOTS are new projects that connect existing systems across national borders. In these cases, intermediaries are introduced to connect the existing components. Another reason for the complex PII structure is reuse. Respondent R3.12 of the Vaccination case study noted: *“As all JGZ organisations were now connected to the platform to exchange files with each other, it was a logical phase for the RIVM to connect to the LSP to share data with the JGZ securely”* (R3.12). Here, we observe that new message exchanges connect to pre-existing components.

Additionally, one can observe that these sub-configurations are evolving towards a networked or distributed model. *“So now we have a node model, where you send messages back and forth according to a relay model. Moreover, we see that we want to move towards a network model, (...) where you start working with registries.”* (R3.10). In the Child Authority case, the focus is on a renewal programme to transition the youth and security chain from a relay (or handoff) model to a network model. Both case studies demonstrate a focus on integrating different datasets. *“Collaboration needs to be supported more flexibly, transparently, and supported within the organisation”* (R3.11). The respondent argues that there are three advantages to a network model. Firstly, it helps citizens better match their needs. Second, users gain a deeper understanding of various processes. Lastly, there would be better control of the data. The idea is that, in these cases, the data stays with the source, leaving the source owner responsible for timeliness and access to the data.

Another area of further development within this sub-configuration is eInvoicing. Connecting to the Peppol network is also a step towards a network-based model for eInvoicing. There is a difference in data models here, shifting the emphasis towards data retrieval and notification of changes. The underlying idea here is data at the source. In contrast, a model like Peppol focuses more on enhancing interoperability and harmonising infrastructure.

In exchanges, standards, and agreements are the norm. Conversely, in every case configuration, we observe cooperation between different public parties that jointly purchase services from a public or private intermediary. These central systems involve cooperation among various parties, supported by an intermediary that provides a system. Such systems are supported and utilised by parties with shared interests. For example, in eInvoicing, different central government bodies must be able to receive the same type of invoice. In the case of the WMO, municipalities need to share social assistance information in a similar manner. In each case, the parties involved had similar needs and made comparable requests.

Within these central systems, multiple services can be provided to support the process. The Vaccination respondent 3.12, for example, told us, *“The LSP has several functions. It ensures that parties can only connect using certain standards with qualified systems that have been demonstrably sufficiently tested. This ensures that only healthcare providers are permitted to connect. It also controls authentication based on certificates.”* The process, therefore, includes not only sending and receiving but also services such as system queries, identification and authorisation, data transformation, and (technical) data validation.

The data indicates that the exchange will become more decentralised. Nevertheless, there will still be a need for central services, such as the view service or a registry. In some cases, services have been

established centrally. For example, within OOTS, there is a central register that displays which organisations offer data, the standards for data exchange, and the types of evidence that can be exchanged.

There are additional services offered at many different points within these more complex PII. Across the cases, we observe services such as identification, authorisation, notifications, track-and-trace, portals, data quality checks, access logging, and (trans)formation services. These are more complex forms of data sharing than the data dump in configuration one. It involves constant development to improve the process and gain more insight into the exchanged data. For instance, respondent 3.10 of the case WMO stated, *“You can say something about processing times and waiting lists based on those messages. Nevertheless, you can also comment on the payment behaviour of municipalities to an individual provider. So, you can extract information from that data.”* Much of this is still in its infancy.

Within each of these exchanges, there is a growing emphasis on the use of domain-specific standards. *“Data exchange through messaging must comply with iStandards.”* (R3.10). Here, authorisation and identification mostly use EIDAS and national e-ID facilities. The e-delivery network is also used within various data exchange systems. *“This is a messaging-based platform for message exchange”* (R3.21). For example, a respondent from the Youth & Safety department stated: *“We have always used Digilevering EBMS messaging. For a year or so, Digikoppeling has also included APIs as an option. In addition, the VNG is developing a notification standard.”* (R3.11). We have noticed that public standards are often utilised in this technical sub-configuration.

6.2.4 TECHNICAL SUB-CONFIGURATION 4: DISTRIBUTED PII

Technical sub-configuration four is characterised by the distributed nature of the infrastructure. There is a many-to-many relationship within the network here. It involves collaboration with both public and private intermediaries that provide services to senders and receivers within the network. Open standards are the norm, and a lot of technology is reused. Both push and pull mechanisms are possible in this configuration. The Kik-V, Edu-V, and Carecodex cases can be characterised as distributed infrastructures. As respondent R3.20 noted: *“You never solve the problem if you only solve it as this kind of referral, passing from A to B, then B to C, then C to D, and so on. It is rarely one logical chain of links coming one after the other. So, we don't talk about data exchange chains, we always talk about networks instead of chains”*. Rather than chains, cases involved in this sub-configuration tend to be networks.

A trust framework is central in this sub-configuration. The agreements within this trust framework are implemented individually by the involved parties. Typically, this is done through intermediaries, but the parties can also develop and implement them independently. As the Kik-v respondent mentioned, each sender and receiver is responsible for connecting to the network. The specifications are incorporated in the trust framework. This standardisation allows parties to prepare for connection and data exchange with the entire network. Such uniformity makes adoption easier for the parties. *“The chain parties have therefore set up a way of exchanging data: sending a question and a reply back and forth with certain standards and infrastructure. As a healthcare provider, you can prepare for this too, because every party connected to Kik-V will ask you to set it up in the same way”*(R3.15). Collaboration and the use of (open) standards play a significant role here.

Collaboration between different parties is sought to prevent the endless repetition of work and agreements that each party has to carry out and make, respectively, during the exchange. However, in the beginning of innovation, repetition of work cannot always be prevented, as respondent 3.20 noted: *“You can do it in one region and think, ‘Well, then we have it all worked out.’ Then other regions can adopt it. But no, every ICT department of every hospital starts from scratch. Everyone wants to make sure it is right for them”*. In this case, a policymaker united the chain partners and their needs and began seeking common ground. Cooperation was pursued voluntarily to strengthen the network.

Sometimes, this collaboration is outlined in a covenant. *“So Kik-v is an agreement set, or a set of agreements. It is concluded by several parties, who have therefore signed a covenant among themselves. That covenant basically says, ‘We will work together and abide by the following agreement.’”* (R3.15). At Edu-V, they noted at the outset that not all parties were eager to join immediately. *“In the run-up to the agreements, one party said, ‘I do not need it at all; I have my own ecosystem’. That is a problem because that party plays a key role in the system. In the end, with some reluctance, they agreed to participate. We will see how it goes. The great thing is what you can see now because we are just taking those phases now, and they are thinking, ‘Wait a minute, this is really going to happen.’ So that party is suddenly taking a very active stance.”* (R3.8). Ultimately, this results in intensive cooperation within a network and the involvement of different actors.

In a distributed IT system, reusability and the use of standards are especially important. A respondent 3.8 from Edu-v stated: *“It has to be as low a threshold as possible. This is also important from a competitive point of view, and of course, it is also important that everyone gets an equal level playing field”*. Increased support is achieved through the use of standardised interfaces. This lowers the barrier to entry for small parties and boosts private-sector involvement. It can also help prevent vendor lock-in. *“Well, you can imagine that it was quite a big phase for all parties to conform to that. However, once we had reached an agreement, it generated much support from the smaller parties. This is because the small parties can now test the market much more easily. They can immediately receive standardised data free of charge. These are agreements that provide enormous support. Also, for the smaller parties.”* (R3.8).

This network focuses mainly on technical interoperability and standardisation. The idea is that parties should be able to find each other easily and establish connections. *“It is all about the front doors of organisations. A key phase we are taking is to standardise all of these interfaces. So, when you enter the market and want to participate in the ecosystem later on, you will know exactly which interface you have to comply with. If you implement it, you will immediately have access to all of these other parties’ front doors”* (R3.8). Parties to the Edu-v case also receive a seal of approval indicating compliance with the standards and agreements. This enables other parties to recognise those with whom they can more easily cooperate clearly.

However, this does not mean that message standards should be uniform. *“We are dealing with fifteen source systems. If you wait until all of those source systems have the same standards built in, it will take thirty years. So, we said, just give us the standard you have or the data from the source system, and we will translate it. At one point, work was done on a translation service specifically for a number of systems. That is quite unique. You cannot use that for anything else.”* (R3.20). Intermediaries, especially private ones, are specifically there to format data and offer conversion services. Therefore, specific services such as conversion services are additionally available.

In a distributed network that retrieves data (pull), there is often a registry or lookup function to retrieve the correct data. In such a network, this is usually built into the sender's application. The Viewer function in Care Codex can fetch relevant data at the time of retrieval and display it. Within KIK-V, the vision is that this function will eventually be integrated, but a temporary central module has been designed in the meantime. *“Yes, there is now an intermediary in between, but that is only a temporary solution. That is not necessarily our final vision. Our end goal is to create a kind of federated network of questioners with their own node, who can ask a question to a data station with a node as well (...). This way, everyone can simply ask a question from their own application.”* (R3.15). In this sense, KIK-V is not yet a fully distributed network but is evolving towards it.

Compared to other technical sub-configurations, a key function within a distributed PII sub-configuration is controlling access through identification and authorisation. For example, in the Care Codex, there is a relatively open network. Therefore, it is important to identify connections and verify whether users should be granted access to each node. This point is also supported within Kik-v. Here, the respondent refers to the application of the EIDAS regulation. This regulation outlines concepts and standards for credentials, specifically the digital proofs required to exchange information within a network.

In this sub-configuration, the private sector plays a significant role. The fact that agreements are in place and open standards are used makes it easy for the private sector to join the network. *“One of these frameworks was that the ecosystem surrounding this data exchange should be modular. This means that for every part or modular piece, it is preferable to have more than one provider. This prevents vendor lock-in”* (R3.20). With the agreements in place, private parties can offer systems and services to sending and receiving parties. These systems can be integrated with existing systems or offered as standalone systems on top of existing systems. *“There are now about five or six vendors providing that. They all have their own solution, but with the same functionality”* (R3.15). Nevertheless, finding enough private parties to provide this service can be challenging. *“Ideally, we would have preferred at least three providers of the integration layer. In the end, it became one. We did not like that. That is not a real commercial club, so that makes a difference. They also favour it being open and reusable”* (R3.20). Here, private and public interests may clash, leading to vendor lock-in. Ultimately, private parties with a commercial interest will proceed only if a positive business case can be made and sufficient market growth potential exists.

For example, respondent 3.15 in the KIK-v case describes starting a trust framework and distributed PII as a 'prisoners' dilemma'. Broad adoption requires agreements, but shared agreements require broad support. As long as there are no agreements, parties will not implement them quickly. A critical mass is needed to work together and change the domain.

6.3 GOVERNANCE SUB-CONFIGURATIONS

In this section of Chapter Six, we perform a third hierarchical cluster analysis to uncover the underlying structure of governance within a PII. The term governance refers to the structures in place that distribute roles and responsibilities, providing the means by which (part of) the PII is managed. They encompass: *“formal goals, ownership, distribution of property rights, and decision-making procedure, that fosters and sustains collective*

action” (Enjolras, 2009, p.10). These governance structures typically consist of formalised meetings, regulatory frameworks, sustained funding mechanisms, and clearly defined roles and responsibilities.

The analysis reveals that design choices manifest on two distinct levels. The first is at the level of overall governance configuration in a PII; the second concerns the individual governance structure. To account for this distinction, we have split the analysis into two parts. As not all governance variables were taken into account in this analysis in this section. Some of the governance variables were more closely tied to the individual level within a structure and are therefore included in the analysis of a single governance structure. Therefore, we first analyse the governance configuration at the level of the entire PII. Then, in Chapter 6.4, we zoom in on the specific governance structures that respondents identified as most significant within each case.

To gain insight into the governance structures in a PII, this section examines the governance design variables at the PII level. The clustering is formed by some of the variables from the framework as presented in Chapter 5.3.2:

- The number of governance structures,
- The ordering of the governance structures,
- The scope of the governance structures,
- Multiple levels of governance, with additional focus on the distinction between regional, national, European, and global levels.
- And the involvement of the public and/or private sector.

Together, these variables provide a nuanced picture of how governance configurations can be organised within PII and how coordination is achieved across different actors and levels. As this dataset contains mixed variable types, the Gower distance measure is used. Within the clusters, we found both global patterns and fine-grained subdivisions.

Figure 6-4 shows the dendrogram of governance PII sub-configurations. The dendrogram shows how the governance-related variables are grouped based on their similarity. The figure illustrates six distinct configurations, each marked with different colours. The x-axis represents the cases, while the y-axis represents the distance between them. Cases that are entirely different will have a distance of 1. The smaller the distance between cases, the more likely it is that these cases form a cluster. For example, the cases BRP issue reporting and Notification of property value are very closely related and share the same configuration; therefore, they are levelled in the same cluster. The distance between these two and the case of the National housing process was also small (lower than 0.3), so they fell into the same cluster. However, compared to the OOTS and eCodex cases, the difference is large. The case of OOTS and eCodex resembles no other cases but each other, forming their own cluster. To obtain the final configurations, we use both the dendrogram and a qualitative data analysis. This combination ensures that the cluster not only reflects statistical similarity but is also grounded in the contextual nuances and insights derived from the qualitative interviews.

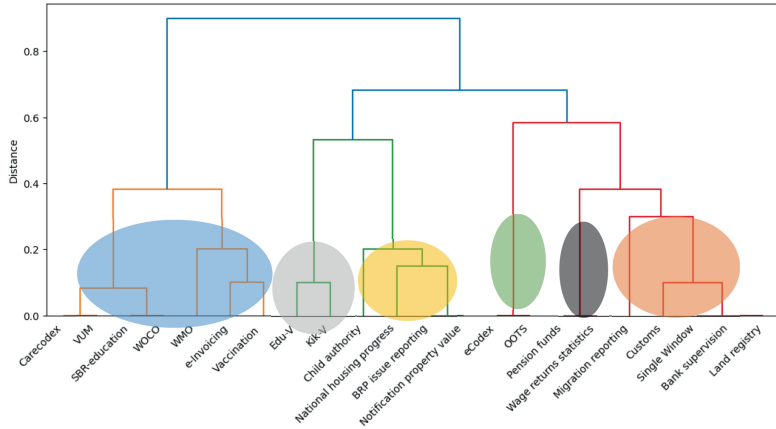


Figure 6-4. Dendrogram of governance variables

Based on the dendrogram analysis and qualitative analysis, we identify six governance sub-configurations. Here, we observe that the sub-configuration marked with the blue circle includes more cases than, for example, the sub-configuration marked with the green or black circle. This shows that some sub-configurations are more widely applied than others. This could indicate they have existed longer and been more widely adopted, or that they were deemed more valuable and applicable.

Table 19 presents an overview of the six governance sub-configurations, the design choices characteristic of each configuration, and the cases that exemplify this type of configuration. Each row represents a different sub-configuration, with its own characteristic set of design choices. In the following subsections, we will examine each sub-configuration, its characteristic design choices, and examples from the cluster.

Table 19. Overview of governance sub-configurations, their characteristic design choices and the representing cases.

#	Governance sub-configuration name	Design choices characteristic of the configuration	Representative cases
1	Central European with a focus on technique	Central governance structure Multi-level: International and National Scope: Technical-focused	OOTS, eCodex
2	European governance with an underlying national structure	Central governance structure Two governance structures, national and European. Scope: Technical (standards) and Data	bank supervision, Import, SW, Land registry, Migration, (eInvoicing)
3	Operational meetings, limited structure	Mostly operational meetings Scope: Data	Pension funds, Wage returns statistics.
4	Different governance for systems and data formats.	Multiple governance structures Collaboration is high Multiple governance structures Scope: Data and technical system	BRP, Notification property value, National housing progress, Child authority
5	Central trust framework	Central governance structure High collaboration Inclusive Scope: Data and technical standards	Edu-v, KIK-v
6	Decentral governance structures (some with additional trust frameworks)	Multiple governance structures Multiple levels Network of hybrid relationships Scope: Technical systems, Technical standards, Data, Data formats, Governance.	VUM, CareCodex, WOCO, SBR-education, eInvoicing, WMO, Vaccination

During the analysis, we found some overlap among these sub-configurations. For example, the eInvoicing case study features both a decentralised governance structure and a European focus, with a Dutch supporting committee. We therefore included it in both sub-configurations during the analysis.

6.3.1 GOVERNANCE SUB-CONFIGURATION 1: CENTRAL EUROPEAN WITH FOCUS ON TECHNIQUE

The first governance sub-configuration is characterised by a central governance structure on a European level. Within this sub-configuration, the focus is on the technical implementation of data exchange, which replaces an existing data exchange. In particular, agreements rather than systems are maintained within the governance. We observe that these are often emerging infrastructures. The two cases that represent this sub-configuration are OOTS and eCodex. *“EU-LISA has a very tight governance structure. You can see this on their website, and you can also see that the Netherlands is represented. Therefore, EU-LISA has a specific governance structure comprising a board and a system of various workgroups. Therefore, whenever we want EU-LISA to take action or EU-LISA requires something from the Netherlands, we are represented in these bodies. (...) It is up to us, the member state itself, to make sure that we stay connected with our national stakeholders to bring the right points there.”* (R3.7). Countries have representation in the European governance structure to express their national perspective. This suggests

that national consultations should be conducted to establish a shared position. *“Yes, so that is what I actually meant when I said that this governance is still very much in its early stages. So, there is a European governance structure, and the Netherlands is part of it. We still have to work on how to make the national lines flow smoothly, so that we provide the right input on a European stage like this”* (R3.7). In both cases, a tightly regulated governance structure exists at the European level; however, there are still limited or no national consultation structures.

Within the working groups, agreements are primarily made in the technical domain, such as focusing on technical standards, security issues, deployment, testing, and the findability of documentation. Data exchange content is less frequently discussed within these groups. Both OOTS and eCodex aim to digitise existing documentation. The respondent from OOTS stated: *“So the next level of the data exchange is that you exchange based on agreed structured documents. Only, we are at least realistic enough to know that it is a very long road, if it takes twenty-seven countries to agree on something like that, it is a very painful road, so that is why we take this approach.”* (R3.21). This shows that it does not imply that this will not be addressed in a future follow-up phase; however, the initial logical phase concentrates on technical aspects.

Often, these exchange processes are based on laws and regulations. The respondent of eCodex stated: *“There have been legislative negotiations in Brussels on how a European arrest warrant should be. Based on that, forms are made, and a lot of those forms you can also find on the European e-justice portal. There, you can find all kinds of standard forms that are used in these kinds of regulations. So, it is determined at the European level what should be in these forms.”* (R3.7). Especially when data exchange occurs across national borders, coordination will take place at the European level.

Both cases show that this construction is a precursor to another configuration, sub-configuration two (or higher). *“So indeed, we need to set up a Governance for that. And that, yes, we are still exploring that. That is still, there is still very limited.”* (R3.7). This will follow the European governance and will slowly work its way through to national alignment as well.

It is important to ascertain the parties involved in the (national) governance structure and in the interpretation of national legislation. *“Then it, the ministry, of course, is the coordinating party for that. And here from the ministry, we try to involve all stakeholders as much as possible, so all organisations that will be involved in the future. So, there is a lot from the ministry looking outwards and trying to engage the stakeholders.”* (R3.7). In this context, it shows that ministries with a mandate and decision-making power have the leading role.

6.3.2 GOVERNANCE SUB-CONFIGURATION 2: EUROPEAN-DRIVEN GOVERNANCE WITH NATIONAL GOVERNANCE

The second governance sub-configuration is characterised by two elements: a central governance at the European level and an additional national governance. The national governance has a hierarchical relationship with the European governance, acting as a parallel governance structure. In this sub-configuration, both data and standards can be discussed. Examples of this type include bank supervision, import, SW, land registry, and migration. Invoicing also involves a European-driven governance to some extent, with a national parallel governance structure, and is therefore considered part of this configuration

to some degree. In the national governance structure, countries develop their own perspectives and implement European legislation.

All cases involve European-level collaboration on the content of data exchange. In some cases, it includes data formatting, while in other cases, there is still room for national implementation. Often, this depends on whether the data is to be sent to a European organisation for monitoring. Both the cases of migration and bank supervision require reporting to the European Commission. In the case of bank supervision, *“For banks, they call it CRDV4. Those are statutory reports that they have to prepare and submit to DNB, where they have fulfilled the statutory reporting obligation. We, as DNB, then sometimes have the obligation to forward those to the European regulators. When data exchange is done by banks, this will be the European Bank or the European Banking Authority. When data exchange is done by insurers, this is EOPA. Those are European institutions to which we send our reporting. Sometimes 1-to-1 and sometimes we do some editing on top of that.”* (R3.4). In both cases, the national receiving organisation may require additional data from the data supplier. In the bank supervision case, based on European and Dutch legislation, banks and insurers are obligated to send their (financial) reports to the DNB. In doing so, there is an additional consultation structure in place at the national level between the DNB and the submitting parties or auditors. They also cooperate with CBS, which receives (part of) the data.

The European governance in all cases is part of the European Commission or is part of central organisations funded by the European Commission. Customs declaration involves DG Taxud, SW, and eInvoicing involves DG Grow, and Migration involves Eurostat. The Directorate-General for Taxation and Customs Union (DG TAXUD), Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (GROW), and Eurostat are all part of the European Commission. They create rules and regulations on a European level for countries to implement. During implementation, national collaboration is sought for data exchange. In the case of the Land registry, the European Land Registry Network (ELRN) is involved, which is funded by the European Committee. Interestingly, eInvoice also has, besides the European CEN TC434, a trust framework called Peppol. This once started as a European project and is now an international governance effort in which parts of the infrastructure are being collaboratively developed worldwide.

At the national level, there is a parallel governance structure that provides input to European governance. As respondent 3.19 of the case of Migration noticed: *“Of course, there are all kinds of consultations. The government works like this: before a Director-General, or even a minister, has to make a choice, there are already many consultations prepared for that choice. In our case, (...) we have a consultation structure with the chain partners, if we want something”*. Additionally, programs can be started due to new developments. As respondent 3.22 of the Single Window case stated: *“We agreed that if there are major changes and large developments on the European level, we can create a project or program. That is also what happened with the European Maritime Single Window. The responsible ministry can then coordinate this.”* In parallel with the governance structure of the European work group, operational meetings are held to implement the various (European and national) rules and regulations. *“We have consultations with the professional body, so that is the high-level group, along with several stakeholders, network notaries, several premia collecting offices, so to speak. Then we have consultations with all seven software vendors. Those who sit at the table*

with our IT organisation, because that is very convenient, speak to each other better than we do. Then we have another group. If we want something with a new style sheet, we ask notaries to come here and help us make a text, because it is your text, not ours. So, we have consultations, not in one joint session, because at a Polish country day, software vendors, notaries.... So, you have to pull that apart by topic, say, technical or legal content or efficacy, say, you have to pull it apart.” (R3.13). This mainly concerns the ins and outs of the day-to-day business. Depending on the consultation's content, appropriate stakeholders are invited to the table to discuss topics such as data, systems, and standards.

In the Single Window case, one of the receiving parties also serves as an intermediary. As a result, there are four layers of technical management and customer consultations between the receiving parties.

“If parties want something, so technically functional in terms of data changes, they raise it at one of the consultations. And then we engage with the parties about what this means. What needs to be changed for it? Can that be done? Do they want that? Is that allowed? Do the parties want that? What about costs?” (R3.22). Here, the recipients talk about the technology and content of the data exchange.

In all these cases, we observe that there is some influence by the private sector. As respondent 3.8 of the case of import declarations stated: *“During this phase, the market does not have decision-making authority as the regulations from the EU have to be implemented, but they still have a lot of influence in the national implementation process. As every country can apply the regulations within their context.”* The private sector has no decision-making rights but can influence the decision-making process. And even though they lack decision-making rights, there are still many implementation choices to be made in the process. *“Companies have to change their systems. Let me say that the format of the message changes, or the way they must report changes. Then they have to modify their systems themselves. There is no compensation for that, and we just wish them good luck. We do try to shape it so that they have to adapt as little as possible. That is responding to business needs. But in the end, they just have to do it. It is legislation”*(R3.22). It is by joint effort that the public and private sectors seek solutions.

The organisation responsible for the land registry provides tooling and style sheets to software vendors, who, in turn, are free to implement them and offer them to their customers at their own time. A distinction can be drawn between the Land Registry case and the cases of SW and Import Declaration. The latter falls under the jurisdiction of the European Commission, and from a certain point onwards, all parties must comply with these laws and regulations. In the Land registry case, the provision of digital services via XML remains voluntary, with tooling and stylesheets optional extras. However, at the end of the day, *“Software vendors also have to invest. They themselves have to finance all these modifications to their system, all by themselves. So yes, that also gives a certain choice in what you make and do not make.”* (R3.14).

Despite cooperation between public and private parties, the requesting government party has the final say on the national level. However, many private sector organisations are represented in the European decision-making process. *“The union of loaders, the union of Customs brokers. They all have a representative in Brussels. From these consultations, a decision is taken at some point, and then the Netherlands has to implement it.”*(R3.9). Therefore, having a national representation from the government's perspective is also important. At first glance, governments seem to show perseverance in decision-making. However, closer inspection shows that the private sector, especially through collaboration in (European) working groups and lobbying efforts, is becoming increasingly influential in the decision-making process.

6.3.3 GOVERNANCE SUB-CONFIGURATION 3: OPERATIONAL ALIGNMENT

In governance sub-configuration three, the governance structure of the PII is minimalistic. The PII is characterised by a limited number of parties. Additionally, there is only national collaboration within the chain, and most of the current consultation is operational. Nonetheless, respondents suggest that establishing a broader governance structure to cover various layers could be beneficial, as it would help and inform those involved.

In the case of pension funds, for instance, there is limited governance for this issue. The pension administrators have a higher-level committee with representatives from the directors, where they discuss several issues, including data exchange. Here, the scope is much broader than just the use of UWV retirement data. These are mainly ad hoc discussions, though the respondent indicated that they are looking at a more dedicated governance structure, given that, in the current situation, it is difficult to raise and resolve issues quickly and in an integrated way. *“A dedicated collaboration structure, whether you call it governance or not, would actually be very desirable. A chain-wide, with ministries and other parties, not only the pension funds. With the current problems, we are getting there.”* (R3.1). Although some PII users seek a more defined governance structure, collaboration remains limited without the UWV's priority or resources.

In both cases, it is the supplying party that must coordinate with one or a few receiving parties. This happens uniformly from the senders' perspective, so the alignment around the system is less than in other configurations. *“Governance is basically not necessarily about how that data is delivered. That is more about the content of the policy. And maybe if you are going to set that up very differently in the Netherlands, that data sharing, you will then hang that transport in such governance”* (R3.18). We observe that central processes characterise this configuration. There are also tactical and operational consultations, but they are centrally arranged. Here, the focus is mainly on the content of the data.

6.3.4 GOVERNANCE SUB-CONFIGURATION 4: GOVERNANCE FOR TECHNIQUE AND DATA CONTENT

Governance sub-configuration four is characterised by the separate governance structures for data, for systems, and for standards. There is no clear hierarchy between these governance structures. However, systems must follow the prescribed data formats as set out by the data governance framework.

For instance, Digimelden has a governance structure for connecting standards and a joint stakeholder meeting for the Digimelden system in the case of BRP issue reporting. The provision of both systems, standards and governance is the responsibility of a public intermediary. This way, the facilitation of collaborative decision-making and cooperation is ensured. In the case of the Child Authority case, a central system provides Customs insight into youth cases, including national person registration (BRP), rulings from the judiciary, and youth healthcare organisations. Each participating organisation, both senders and receivers, has a contract with the public intermediary, and together they collaborate on the services provided. Due to the increasing needs of the youth care domain, an additional technical governance has been

established, of which the service for insight into custody is part. As respondent R3.11 stated, they started a collaboration to build multiple systems and services that can be used for several organisations in youth care.

In other cases, each system or service has its own governance, for example, the case of issue reporting (BRP), each system has its own governance structure. *“If you look, for example, at the situation with the RIVG, the basic registration of persons, they have periodic consultations where they actually look at the efficiency of their reporting process”* (R3.1). Respondent 3.15 elaborates that it is about more than just the day-to-day coordination. It is also the coordination of further developments, prioritisation of issues, and discussion of current needs.

In a governance for a system or service, three layers, operational, tactical, and strategic, are observed. *“Operational are the Ministry of Justice and Security, the RINIS Foundation, and all the intermediaries of the municipality. For example, CGI, PINK-Rocades, etc. At the tactical consultations, you talk more about the director of the tax department, the UWV, and the SVB. And strategically, the directors of those organisations, the state secretary and ministers participate.”* (R3.25). The tactical consultations focus on the standards used within the facilities. *“That tactical consultation is really about the standards used. It is not about Digilevering, because those standards are separate from it. Digilevering is now the current vehicle through which those standards are enforced. It is actually about the Digilevering standards that lie beneath it. That is what needs to be decided on.”* (R3.25). At the strategic level, you often find a high level of alignment with clients and stakeholders. *“We have coordination with IPO on this project. That is the umbrella organisation of the provinces. That stands for Inter-Provincial Consultation.”* (R3.16).

Data content and data exchange requirements are discussed at different levels. Although there is separate governance for data and systems, a hierarchical relationship exists. Standards and data formats are often developed in separate working groups, and systems must comply with and implement these standards and formats. For example, in the case of BRP, the notification service is part of the personal data registry system. A customer using the personal data registry should be able to report issues and errors in the registry. Consequently, the data format depends on the registry's format. The governance of the notification service is therefore subordinate to the core registries. Additionally, many government services must comply with architectural models established at higher levels. For instance, in the case of property value notifications, the municipal model architecture (GEMMA) provides rules, guidelines, and data formats for exchanging property value data. Regarding National housing progress, organisations such as VNG, IPO, and BZK, which operate at higher organisational levels, develop collaborative agreements on data formats, privacy security, and data usage by the recipient. These requirements must be implemented and considered by (intermediary) organisations developing systems and software for data exchange.

6.3.5 GOVERNANCE SUB-CONFIGURATION 5: CENTRAL TRUST FRAMEWORK

The fifth governance sub-configuration is characterised by a central trust framework for data exchange. This framework enables central agreements within the domain as much as possible, while also relying on bilateral implementation and coordination. In this particular sub-configuration, a substantial number of market participants engage in the governance structure. This sub-configuration appears to originate from a programme, which is defined as a short-term, temporarily funded project prioritising innovation and

progress. This allows for the exploration of new structures. Once established, there is only national collaboration, with no direct European cooperation. The trust framework can make agreements on data formats and technical standards; however, it does not focus on providing systems or services directly. Operational meetings occur solely between system or software providers and customers; this is not part of the trust framework. Within the trust framework, different meetings are held to discuss various aspects of data exchange.

Important in a trust framework are the collaborative nature and the reliance on agreements. *“Directors of all parties involved who have signed a covenant with each other. So Kik-V is a deal package. (...) And that covenant basically says, ‘we work together, and, in this cooperation, we keep to the following agreement.’ So that agreement set is maintained and managed by the management organisation. But under the responsibility of those administrators in the Council.”* (R3.15).

In both cases, there is an extensive range of working groups. For instance, there is an architecture board, and there are six additional working groups where substantive discussions take place on different topics. On top of this, there is a sounding board and focus groups. The purpose of these last two is to test and gather feedback from various stakeholders, and to allow user groups to have a say (EDU-V). A characteristic of this sub-configuration is a broad representation and thus a large support. Within Kik-v, there are representations of nursing care and healthcare providers. They represent a large part of smaller branch organisations (R3.15).

The principle of a trust framework is that agreements are made about standards and data formats for the data exchange. This means that no central services are provided, but every organisation can implement the standards in its own systems or use a private-sector solution. In one trust framework, there are temporary central queries. It is created to demonstrate the applicability and increase adoption of the standards by requesting organisations. This is intended to be temporary, however, and will eventually be replaced by market functionality.

Focusing solely on setting standards without developing services provides a business opportunity for the private sector, as they play a significant role in the trust framework for designing and managing these standards. We highlight that a trust framework is often one of the most inclusive governance structures.

6.3.6 GOVERNANCE SUB-CONFIGURATION 6: DECENTRALISED GOVERNANCE

Governance sub-configuration six is characterised by decentralised governance structures. These can include various forms of governance structures such as temporary programmes, trust frameworks, systems governance, data format governance, standards governance, alignment meetings with the private sector, and (bilateral) operational meetings. In some cases, we observe European or international collaboration, but in every case, there is a national governance structure. Most of these are decentralised; however, some have a hierarchical relationship. Many governance structures in these configurations extend beyond the boundaries of the original data exchange and focus on harmonisation across multiple domains. Within this configuration, it is possible for multiple trust frameworks to be involved. For example, in the case of Carecodex, there is a trust framework for BabyConnect and a trust framework for MedMij (R 3.20). The

first defines agreements specifically related to birth care, while MedMij establishes agreements for the exchange of personal health data.

To some extent, there is centralisation, as there are specific governance structures for data formats, systems, and standards. The respondent of eInvoicing noted that there is a central governance for public-private standards, one for public-private data formats, and a central governmental governance structure for public sector collaboration. In some cases, external third parties such as the standardisation organisation NEN are involved and responsible for organising collaboration between the public and private sectors. *“Yes, there are even NEN standards for that, so that every party involved can take a role in that governance. That is the NEN 7522. A NEN committee that deals with the governance of information standards.”* (R3.12)

It is not uncommon to observe delegates from one consultation participating in another consultation structure. *“In all consultations, we meet the same partners every time. (...) There should then be directors there, but they did not come. They were represented by exactly the same representatives who do the preparations in the other consultations”* (R3.5).

Additionally, intermediary organisations tend to represent their customers, as they possess more technological expertise on the matter. For example, Logius, which provides a system in the case of WOCO, participates in the SBR-Wonen workgroups and also manages its own governance for the Digipoort system. It can be challenging to ensure the participation of the intended target groups in relevant consultations. If too many parties are involved, progress might stall; however, too few participants can result in lower adoption rates and weaker support for agreements. Trust in these governance structures is crucial. As respondent Carecodex observed, it is difficult if the right people are not involved in decision-making. If they are to face the outcomes, they should be included in the process. The respondent noted that this is one of the hardest aspects of designing governance and is likely a reason many projects fail. In this context, having numerous governance structures and meeting schedules can be counterproductive. *“Well, one of the objections we do hear is, does there have to be another consultative body? We already have so many. (...) We are now going to map them: Where are all the consultations taking place? What is decided in them? There are actually two pieces of data. One is that all those consultation structures are in place. So, adding something is not really logical. But at the same time, it is also true that they are not going well. So, then we have to see what could be changed in those consultations that all take place. Maybe some of them could be taken out!”* (R3.5). Keeping a balance and maintaining an overview is therefore key for the stakeholders in the PII.

The relationship between the different governance structures is sometimes hierarchical, but more often resembles a network. The relationship between the two entities involves mutual influence, yet it is important to recognise that each governance structure operates under its own distinct mandate and set of stakeholders. Respondent eInvoicing noted that these networks may also employ their own standards. For instance, the Peppol network focuses on Peppol standards, while the central government system adheres to its own governmental standards. Depending on your position within the network, you will need to adapt to these different standards. This offers flexibility but also necessitates additional services. Respondent WMO observes that compliance with various agreements is not always well monitored. An important question is: in a public-private partnership with consensus decision-making, who bears ultimate responsibility for compliance, and how can this be enforced? *“They do, of course, always try to consult by consensus. But suppose there*

were a conflict, then a certain voting ratio applies. Should that then fail, then your system holder, so in this case VWS, the Ministry of Health, Welfare and Sport, is the one who can make a decision on that.” (R3.10) Here, it is evident that consensus often serves as the starting point. However, in practice, there is always a final responsible party, especially when it comes to data sharing for public purposes.

6.4 ZOOMING IN ON A SINGLE GOVERNANCE STRUCTURE WITHIN A GOVERNANCE SUB-CONFIGURATION

During the analysis of the governance sub-configurations at the level of the PII, we identified multiple possible forms of governance structures. These include operational governance bodies, technical governance for systems, governance focused on data formats, trust frameworks, and innovation-driven programmes. We pinpointed these structures and, using variables at the level of each governance structure, we can further elaborate on them. As outlined in Chapter Five, previous research has highlighted a set of variables particularly relevant to understanding the individual governance structure of PII. These include:

- The scope addressed by the governance structure.
- The decision-making process.
- The role of the private sector.
- The degree of inclusiveness.
- The financing model of the governance structure.
- The types of participating organisations.

For each of these governance structures, we examine how the aforementioned design variables are implemented in practice.

6.4.1 OPERATIONAL MEETINGS

In many cases, operational meetings are arranged around the chain. These typically involve error-resolution discussions between senders, receivers, and providers of intermediary systems. Sometimes, such as during migration, cases are discussed among multiple parties within the chain. Respondent 3.19, notices: *“Generally, the consultation on data quality is a meeting which just solves practical problems, and signals of errors.”* These meetings are common when individuals are central. The consequences of errors and delays are significant, leading to chain-wide consultations to ensure the process lines are continually monitored and optimised. This shows that communication lines are short and highly valued (R3.19).

On the other hand, it does not have to be centrally organised. Within VUM, for example, this happens regionally. *“There are thirty-five labour markets in the Netherlands. Thirty-five labour market regions. In which municipalities and UWV work together with several parties to ensure that everyone can participate.”* (R3.23). We observe this happens mainly between public parties, and in cases where the number of supplying and receiving parties is low. Though it was mentioned explicitly, it often is not a separate governance structure but a part of an existing governance structure.

6.4.2 GOVERNANCE STRUCTURE FOR SYSTEMS

In many cases, intermediaries develop a central system used by different organisations. As a result, a governance structure specific to this system can be identified. This leads to a governance structure that is used for multiple data exchange processes. Examples include the VUM case, which uses GeVS; three cases (SBR-education, WOCO, eInvoice) that utilise Digipoort as a central system to support electronic messaging to (central) governments; and two facilities within the WMO case. Each of these systems is governed by its own structure, designed to ensure effective management and continuous development. For such governance, system users contribute by paying for the system they use, including financial contributions to the governance's management. These are mainly public collaborations, and decision-making is generally consensus-based. To some extent, intermediating organisations providing these systems can adapt to additional requirements. As respondent eInvoice observed, as long as the functionality remains applicable for their own data exchange, customers of these systems do not mind if another party has extra requirements.

6.4.3 GOVERNANCE STRUCTURE FOR DATA FORMATS

There is a variety of bodies developing data standards. For instance, within Carecodex, Vaccination data case, WMO, and eInvoicing, there is a separate party responsible for aligning message standards. Within eInvoicing, there are even several message standards with their own governance at the national, European, and international levels. It does not have to be the definition of a data format in terms of the technical standard. In the case of migration, a governance structure has recently been established within the justice and security domain for the purposes of definitions. *“It is foremost about definitions. (...) The biggest challenges are always the definitions. One really just has to be clear about this. This is important in any business. I have done that commercially, for example, which is always the challenge, standardising your definitions. (...) Then they are not going to think at all about whether that should be in format A, B or C. It is mainly about how you capture the meta information. That's where they are looking first.”* (R3.19). We notice that structures for data formats can be specified at various levels, allowing for higher-level harmonisation depending on the context.

Currently, all organisations from the Ministry of Justice and their information chain partners are included in this consultation. This leads to well-supported agreements. In many cases, data formats are part of a larger semantic model. For example, in the case of SBR-education and WOCO, a governance structure is responsible for developing taxonomies that support business reporting exchange. Based on this overarching taxonomy, SBR-education and WOCO can identify their own message types. This means that, at a higher level, there is collaboration on definitions, and at a lower level, cooperation to determine which data format will be used. The latter involves discussions between senders and receivers, whereas the former involves receiving organisations only. Many governance structures for data formats are regulated by law or policymakers, but their implementation is a collaborative effort between public and, in some cases, private organisations. These collaborations are often centrally funded, and participation is free for private organisations.

6.4.4 TRUST FRAMEWORKS

In previous chapters, we already identified that trust frameworks are becoming a more prominent governance structure in data exchange with public organisations. Respondent 3.7 stated, *“I think ECodex is a trust framework, especially from the Eu-LISA perspective. As far as I am concerned, they do not call it that themselves, but if you ask me, it is indeed a trust framework. Because Eu-lisa runs ECodex, but there is no central hub. There is also no central registry. They actually only support peer-to-peer connections in the member state. Basically, they set the rules of the game, and they have the supporting software to make those lines work.”* In this domain, the focus is not on system management, but on establishing agreements that the entire domain and data exchange chain must comply with.

Within a trust framework, it is common for work groups to be organised around various topics. For example, within the SBR framework, which includes WOCO and SBR-education, there are working groups specialising in taxonomies (semantics) and groups focusing on technology, such as Signing Technology. The scope of a trust framework is often defined by policymakers or financial contributors. Many trust frameworks are funded centrally or by a select few, making participation in the working groups free.

The trust framework often extends beyond a single data exchange. For example, SBR includes several data exchanges around business reporting, the Peppol trust framework covers not only invoicing but also other (procurement) messages, and AORTA covers the entire healthcare exchange. However, it is worth noting that the subjects discussed are not limited to technology and data content. In the case of VUM, for instance, issues such as fairness are also addressed. *“And part is, for example, in that issue of how much profile I can supply. So, to keep it fairer. Because one municipality of Amsterdam has 40,000 profiles, and the municipality around it only has ten. And you want there to be a fair opportunity. (...) Well, that is where we agreed. You can provide a maximum of so many candidates on a question.”* (R3.23). The trust framework within the Vaccination data case places additional emphasis on privacy in the context of medical personal data. It is therefore dependent upon the specific trust framework as to what is and is not agreed upon.

Again, we observe that representation is important within the trust framework. The trust framework tends to be inclusive and have consensus-based decision-making. *“There are eight roles that are all actors in that (framework). That is the financier or accountable, so the owner of a standard and trust framework. That is the technical administrator, an authoriser, a functional administrator, a distributor, the suppliers, and another expert. So that could be technical experts, so there are all kinds of experts”* (R3.20). It depends on the scope of the framework if private organisations are participating.

There is a growing trend to represent not only technical people but also the direct end-users and stakeholders in the case. Within the VUM, there are working groups with trade unions, agency representatives, and job seekers to represent the direct stakeholders. *“We find that it becomes easier for those organisations if the end-users are well represented and are at the table, because they just tell us, this is what we need functionally.”* (R3.20). The end-users are also represented within VUM and Care codex to bring a broad perspective to the different agreements (VUM, Carecodex). *“We often really lack the healthcare providers themselves who make time*

or have the knowledge to talk about this. Then it is often solved by technical organisation, but then it still does not always work. At least not the way people would like it to." (R3.12) Ensuring representation at consultations can be challenging.

6.4.5 (INNOVATION) PROGRAM

In several of the cases we saw, innovation programs were created. Examples include VUM, SBR-education and Carecodex. These programs are temporary and centrally funded by a policymaker or by a select few stakeholders involved in the case. The programmes are often wide-ranging, with many partners and alignment in the chain. An example is the SBR-education case, where they have recently started to involve software suppliers within the programme, in addition to schools and accountants.

In all three cases, the project's goal is to enhance the current data exchange. In the case of VUM, the programme initiated a trust framework. This involved issues such as data profiles (data format) and harmonisation for data exchange. *"If you want something with your services and you want to modernise or improve it, there really isn't a nicer boost function than a program."*(R3.23). In the case of SBR-education, the programme focuses on improving cooperation, shifting responsibilities, and improving data quality. *"Previously, it was that the government dictated what should be done. The schools or the implementing organisations have to listen to what is mandated. They had complained about that. But they do it because otherwise nothing happens at all. (..) We ask, they deliver, and they do it exclusively because they have to. It was only provided what exactly meets the minimum requirement. (..) That has kind of been the model in which we have operated, say, since the 1960s. And now we say no, we have a shared responsibility. We also have shared interests. We do form a chain, but the one chain partner is not really more important than the other. With one exception, the minister decides what happens. How that happens, you can talk to each other about that. And that is quite a different approach."* (R3.5). Here, one can observe that programmes are also focused on improving cooperation and governance structure in the domain. *"And the other big issue is working together in the chain. How do you do that with each other? How do you arrange that? And when the programme is finished, how do you perpetuate that? Everyone is enthusiastic, but how do you maintain that enthusiasm? And then how do we arrange that with each other? So that is a little thing."*(R3.6) In addition, a programme can specifically aim to formalise existing cooperation in a trust framework. *"We are now in the middle of a programme in the Netherlands to actually formalise those collaborations because there are a lot of requirements there. As an obstetric alliance consists of a hospital and everything around it that does something with birth care"* (R3.20). It can therefore be stated that programs are especially useful for the further development of the PII.

6.5 PREVAILING PII CONFIGURATIONS

In previous sections, we identified the sub-configuration of a PII, including data sub-configurations, technical sub-configurations and governance sub-configurations. This section focuses on answering the research question: **What are the prevailing configurations of PII through which organisations exchange data with (semi-)public organisations?**

A prevailing configuration is a pattern of design choices that either occurs frequently across cases or deviates from other configurations in an analytically relevant way. Using the sub-configurations, we created an

overview of the case and compared the configurations that exist out of sub-configurations. Here, we compared the cases for recurring patterns of sub-configurations that lead to the prevailing configurations. The analysis focuses on identifying recurring configurations that emerge when the entire PII is taken into account. In our study, we review, for each case, the specific data, governance, and technical sub-configuration. We then compare these sub-configurations across cases to identify patterns in PII configurations.

Table 20 presents an overview of the six configurations.

Table 20. The six prevailing PII configurations and their underlying sub-configurations

<i>Prevailing configuration</i>	<i>Data sub-configuration</i>	<i>Technical sub-configuration</i>	<i>Governance sub-configuration</i>	<i>Representative cases</i>
<i>1. Central data reuse</i>	Little structure or some public collaboration (conf. 2,4)	Central infrastructure with minimal collaboration (conf. 1)	Operational meetings (conf. 3)	Wage return statistics, Pension funds
<i>2. EU-regulation</i>	More structured (conf 6,5,2)	Central infrastructure with minimal collaboration (conf. 1)	European governance with an underlying national governance structure (conf. 2)	Bank supervision, Migration, Imports, SW, (land registry)
<i>3. Technical driven</i>	Low focus on structure (conf. 1, 5)	Hybrid infrastructure (conf. 3)	Central European with focus on technique (conf 1.)	OOTS, eCodex
<i>4. Public Collaboration</i>	Public structure collaboration and reuse (conf. 4)	One central infrastructure (conf. 2)	Different governance for systems and data formats (conf. 4)	National housing progress, BRP issue reporting, notification of property value, (child authority)
<i>5. Mixed collaboration</i>	Fully structured collaboration (conf. 6)	Central, or hybrid infrastructure (conf. 2, 3, 4)	Decentral governance structure (conf. 6)	SBR-education, WOCO, Vaccination, WMO, eInvoicing
<i>6. Distributed with a Trust framework.</i>	Focus on content (conf. 3)	Distributed infrastructure (conf. 4)	Central trust framework (conf. 5)	Kik-V, Edu-v, VUM, care codex

We conclude that there are six prevailing configurations, illustrative of how these PII have been designed in practice. Each configuration is addressed in a dedicated section, with a focus on the characteristics and life-cycle phase of the PII.

6.5.1 CONFIGURATION 1: CENTRAL DATA REUSE PII

PII configuration one is characterised by a centrally organised system that involves minimal cooperation between involved parties. There is a single authority installed for the assurance of the data exchange process,

and its related PII quality. This configuration is evident in cases such as the Payroll Tax chain and data exchange with pension funds.

A key feature of this arrangement is that data exchange is initiated at the national level through regulation. Organisations receiving data need it to perform their statutory duties, and this data is obtained directly from central government agencies. In both cases, existing data formats are reused, with the receiving party having limited influence on the exchange process or the structure of the dataset. In these cases, the infrastructure is established unilaterally by the providing parties. The system includes a limited number of services, with no private intermediaries involved in the design or configuration of the infrastructure.

The coordination among the involved parties is mainly bilateral and limited in scope. The data exchange meets the functional needs for information, but its design is limited and of low complexity. This simplicity and centralisation make it an efficient yet not very adaptable model of data sharing.

6.5.2 CONFIGURATION 2: EU-REGULATED PII

Configuration two closely resembles Configuration 1, as both feature a centrally organised system with limited cooperation among the relevant actors. Here, too, we observe a single authority installed for the assurance of the data exchange process, and its related PII quality. However, the key difference in this configuration is that data exchange is initiated at the European level rather than the national level, typically through EU regulations or directives. In this context, the Netherlands has established so-called shadow committees to supervise the national implementation and impact of these European frameworks.

Despite the European initiative, central systems in this configuration remain highly unilaterally organised. The lead in designing the PII usually lies with either the sending or receiving party, with no structural collaboration. European regulatory frameworks primarily govern the data formats. The parties involved possess only a limited degree of influence on the design of the process or the composition of the data format. Participation in the European working groups is therefore essential to exert some influence on decision-making and the standardisation process.

Bilateral and operational processes characterise the coordination of national-level activities. While data exchange is functionally sufficient, this configuration is also characterised by low complexity and limited adaptivity. The emphasis is primarily on compliance with external obligations rather than on cooperation or collaborative optimisation of the PII.

6.5.3 CONFIGURATION 3: TECHNICAL-DRIVEN PII

Unlike the previous configurations, configuration three originates from European initiatives that are primarily technical. Where Configuration two focuses on policy and legal frameworks, these initiatives focus on establishing technical exchange mechanisms for cross-border data exchange.

This configuration of the PII systems is characterised by a higher degree of complexity. The infrastructure involves multiple links in the chain and both public and private intermediaries. Here we identify that multiple authorities are involved in shaping and establishing the PII and ensuring the PII

quality, though these are often within one structure and collaboration, often initiated at the European level. This creates a constellation in which responsibilities and roles are spread across different actors. Additionally, centralisation occurs at the national level. In both cases, Dutch governmental organisations collaborate to create a single node at the national level. Cases within this configuration show that implementation is still at a relatively early stage. The data exchange is functionally available, but the underlying data standards remain only partially aligned. In practice, this often involves exchanging unstructured documents (such as PDFs) or minimally converting existing documentation to XML. Currently, there is no semantic harmonisation or interoperability at the content level. Therefore, the focus at this phase is on building the 'highway': establishing the technical infrastructure and encouraging adoption among participating parties.

Since the data flows in this configuration are generally cross-country, aligning data formats and meanings is challenging. Therefore, the initial focus is on standardising exchange structures such as transport protocols, authentication, and addressing, aiming for future content alignment. In this context, confirming the sender's identity is often more crucial than the immediate quality or consistency of the exchanged data. As a result, the priority is on source trustworthiness rather than data quality or context.

6.5.4 CONFIGURATION 4: PUBLIC COLLABORATION PII

Configuration four is characterised by centralised cooperation in the field of digital systems, where government-wide standards play a central role. This configuration focuses on developing and applying public standards, which are widely (re)used within the public sector. In recent years, there has been a clear shift towards greater market involvement and an increasing focus on market standards. Nonetheless, the basis of this configuration remains primarily rooted in government-developed and managed standards.

Here, we identify that multiple authorities are involved in shaping and establishing the PII and ensuring the PII quality. These authorities are not united in a single structure. We identify a division into two distinct governance structures: one for the systems themselves, and another for the data formats used. The governance of the systems typically falls to public intermediaries responsible for the operational infrastructure and its management. Conversely, the governance of the data formats is held by the data source holders, or organisations authorised to request or enforce data exchange. These separate governance structures are not integrated and are developed at different levels within the PII.

Exchange within this configuration is based on a single data format that is relatively simple. Its strength is in fostering collaboration, encouraging reuse of existing systems, and standardising both processes and systems. By combining a central infrastructure with widely accepted standards, this configuration provides a scalable and reproducible model for data exchange in the public sector.

6.5.5 CONFIGURATION 5: MIXED COLLABORATION PII

Configuration five is characterised by a hybrid PII that combines centralised systems with decentralised, distributed components. This configuration exhibits a high degree of complexity, featuring multiple links in

the chain that involve both public and private intermediaries. The infrastructure supports a wide range of services, such as audit reports using hash functions, technical checks on data formats, identification systems, white and blacklist services for authorisation, and data transformations. Despite this technological sophistication, manual data entry through portals still occurs at certain points in the chain. This offers user convenience for smaller parties with fewer messages. The aim is to encourage these parties to participate in the message exchange.

A key feature of this configuration is the strong emphasis on collaboration in managing shared systems, processes, and standards. We observe the development of governance structures that extend beyond individual information exchange processes, such as the SBR trust framework. There is a relatively extensive integration between systems, standards, and data formats, with actors frequently participating in multiple governance structures. This results in a certain degree of content coherence, even though the governance structures are distributed. The topics are discussed across different structures, but the participating actors overlap significantly.

Furthermore, the configuration demonstrates cross-domain harmonisation. Coordination occurs between sectors and across institutions, aiming to align data exchange at a higher level. Although this cooperation is intense, it is notable that the supplying parties are seldom directly at the table. Their interests are typically represented by private intermediaries who act as delegated representatives within the governance process.

This configuration thus represents a complex but ambitious data exchange model, in which collaboration, integration and technological support are interconnected. These configurations have reached relatively mature stages. The actors and stakeholders are known, and development is ongoing in specific areas within the PII. There are programmes to further develop parts of the PII, but not to establish a completely new PII structure. At the same time, the challenge of further structuring governance and guaranteeing the direct participation of all stakeholders remains.

6.5.6 CONFIGURATION 6: DISTRIBUTED PII WITH TRUST FRAMEWORKS.

Configuration six is characterised by the distributed nature of the PII and a strong emphasis on trust frameworks as a coordination mechanism. In this configuration, the number of central systems remains minimal. Instead, the focus is on developing and managing common standards for data exchange and formats through an explicit governance arrangement. Here, several authorities together share the assurance of the data exchange process and its related PII quality.

Governance within this configuration is centrally organised and occurs within the trust framework itself. These systems are typically designed to be inclusive, allowing various parties, including users, customers, providers, intermediaries, regulators, and policymakers, among others, to participate in decision-making. Decision-making is based on consensus, with mutually agreed arrangements forming the foundation for interoperability within the infrastructure.

A key feature of this configuration is the two-way effect of standardisation. By conforming to the trust frameworks, parties can improve their interoperability with other participants. At the same time, it becomes relatively straightforward to select a suitable intermediary or service provider, as they can show their compliance with shared standards and specifications. However, these benefits only become evident when a critical mass of participants commits. The configuration is thus vulnerable to the classic prisoner's dilemma: without enough participation, the value of the system is not fully realised.

Therefore, the primary focus of this configuration is on establishing effective governance and involving the appropriate stakeholders. The quality and effectiveness of data exchange rely heavily on these coordination efforts. Consequently, configuration six presents a model that emphasises the power of network structures and shared standards, rather than technical integration or central authority.

6.6 QUALITY OF PREVAILING CONFIGURATIONS

In previous sections, we established the prevailing configurations of PII. In these configurations, we establish that there are authorities for designing, maintaining, and adapting the public information infrastructure throughout its various life cycles, to ensure data, technical, and governance quality. In this section, we establish to what extent the prevailing configurations entail data quality, system quality and governance quality—providing future practitioners with an understanding of the impact of a configuration. We tend to answer the sub-research question: **What are the strengths and weaknesses of the prevailing configurations of public information infrastructures in terms of their quality?**

Within this study, we established a quality framework in Chapter 3.3. However, it became apparent that a more extensive study was required to investigate the quality of PII. We chose a simpler method to assess configuration quality, as we still want to provide insights into the quality of the prevailing PII configurations. Therefore, the strengths and weaknesses of the PII were identified in terms of data, technical, and governance quality. From each configuration, two cases were selected for further analysis. The primary respondents from these cases were asked to share the survey with three colleagues within their professional field. These secondary participants were invited to provide their perspectives on the quality of the PII, specifically regarding strengths, weaknesses, opportunities, and threats/risks. Respondents were asked to complete an online survey that included a question assessing the case across three dimensions of quality: data quality, technical quality, and governance quality. No responses to the surveys were received from the technical-driven configuration. However, the interviews conducted in the previous research phases also provided insights into strengths and weaknesses. These findings were incorporated as input for this part of the study and were systematically coded during the analysis process. These results are included in the elaboration of the possible quality characteristics for each configuration. For the technology-driven configuration specifically, only interview-derived items were used to provide insight into its potential quality strengths and weaknesses.

6.6.1 QUALITY IN THE CENTRAL DATA REUSE CONFIGURATION

The Central Data Reuse PII configuration focuses on data reuse. As a result, the data-supplying party often pays less attention to the importance of data exchange, even though it is this party that has developed the systems for sharing. Consequently, the prioritisation of data quality within the chain is relatively low. This is notable, as respondents emphasise that properly organising these message flows can create significant (public) value.

Regarding data quality, respondents generally rated it poorly across all configurations, which poses a significant issue with serious consequences. Because data is reused, an error introduced at the start of the process persists in the calculations performed by all subsequent users. So far, improvements in data quality have not been achieved, mainly due to a lack of prioritisation or knowledge on how to improve, as Respondent 4.13 pointed out regarding the absence of a control framework for evaluating and enhancing data quality. Conversely, Respondent 4.12 mentioned that a new data quality framework is expected to be introduced; however, its implementation and effectiveness still need to be assessed.

Regarding the technical quality within this configuration, respondents observe that it is often outdated. As respondent 4.13 stated: *“The IT processes are lagging behind and are no longer state-of-the-art. There is a technical debt.”* This heightens the risk that it may not be future-proof. The positive aspect of this configuration is that the techniques are proven, have been implemented reliably, and all participants are familiar with the standards and systems, resulting in minimal disruption. This may be due to the advanced development of the cases within this configuration. As in the previous chapter, most cases are in the stagnation and inertia phase. This indicates that the majority of start-up issues have been addressed within the PII, and the participants are familiar with the process.

In the central data reuse configuration, there is little to no shared decision-making. The supplying parties make the decisions, and the receiving parties must accept them. As a result, the quality of governance is perceived as relatively poor. Respondent 4.12 noted that the stewards of the PII have other priorities, which diminish the focus on the PII itself. As a result, cooperation within the PII. As a result of this lack of prioritisation, employees involved in the design have since left the PII. Without ongoing progress and attention, critical knowledge is lost from the PII. However, there are signs that receiving parties and other stakeholders are stepping in to maintain focus on the PII. Additionally, within one configuration, initial steps are being taken towards the next phase, with renewed efforts to foster cooperation, especially at the bilateral level. This development offers opportunities to reconsider and potentially redefine the scope of the PII. Respondent 4.12 emphasises the importance of new coordination and building trust to sustain this progress.

6.6.2 QUALITY IN THE TECHNICAL-DRIVEN CONFIGURATION

In the technology-driven configuration, there is a strong emphasis on technical quality, while data and governance quality receive comparatively little attention. These aspects are often secondary to technical concerns. The cases in this configuration typically involve the development of new PII where existing

documentation is being digitised. The content of exchanged data is sometimes defined by laws and regulations, while at other times, alignment initiatives have been observed but are not yet formalised. The lack of standardised data formats makes it difficult to validate the data. However, according to the respondent from the OOTS case, data quality is believed to improve due to increased confidence in the identity of the sender.

System quality in this configuration can be considered relatively high. Respondents from both the OOTS and e-CODEX cases noted the substantial allocation of resources dedicated to maintaining high technical quality within the PII. Since these are relatively new infrastructures, they benefit from the utilisation of state-of-the-art technologies. This configuration also offers greater technological flexibility. Although central Dutch hubs are available, peer-to-peer integration remains a feasible option, potentially enhancing the overall efficiency of connected systems. In both the OOTS and e-CODEX cases, traceability of the sender and assurance that messages remain unaltered during transfer were identified as essential priorities. Consequently, system integrity and auditability are of notably high quality in these implementations. However, as the respondent of e-Codex observed, less attention has been given to actual system use and strategies to encourage adoption. Whether this will change remains to be seen. A key risk is that a technically robust system is created, but it remains underused. Furthermore, some errors and limitations may only become apparent once the system is rolled out on a large scale. Therefore, a long-term reassessment of this configuration's quality will be necessary.

In the technology-driven configuration, governance primarily operates at both national and European levels. There are various strengths and weaknesses affecting the quality of governance. Firstly, the existence of a European-level decision-making body. Respondents from both the OOTS and e-CODEX cases highlighted that the governance processes are transparent, with clearly defined mandates and structured decision-making procedures. However, the involvement of all EU member states, each with its own legal requirements and regulatory frameworks, leads to slow decision-making.

Secondly, although European governance is established, national governance is still developing. The respondent from e-Codex anticipates a more formalised national governance framework to emerge, particularly concerning the central Dutch system. As governance grows more complex, decision-making may become tougher, and greater effort will be needed to establish high-quality governance.

Finally, European regulation plays a vital role in this configuration. For example, the use of e-Codex is mandated by legislation. As respondent 3.7 of e-Codex explained: *“There is one common denominator between all these exchanges: That is that every exchange must be based on an existing legal instrument or an existing regulation. So, it should never be a question of whether another country is actually allowed to receive the information you send. That is stated in the regulation.”* In this way, inclusivity is legally enforced. Additionally, the technical solutions within this configuration are grounded in EU law, ensuring long-term sustainability. According to the e-CODEX respondent, this legal foundation guarantees that the technical infrastructure will remain in place even after EU funding ends.

6.6.3 QUALITY IN THE EU-REGULATED CONFIGURATION

Starting with data quality in the EU-regulated configuration, the quality of PII data is generally viewed positively. Respondent 4.9 credits this to the use of standardised processes and clear governance structures. Other respondents emphasise validation methods, such as involving accountants and external validation services, as key contributing factors.

Looking ahead, Respondent 4.10 suggests that cross-validation between different datasets should be possible. However, achieving this requires greater alignment of semantics across the various data formats used in different PII systems. Respondent 4.10 suggests that future developments should enable cross-validation between datasets. Nevertheless, this necessitates increased semantic alignment among the diverse data formats, extending the current PII. For instance, in the case of Customs, their PII and that of the Dutch Food and Consumer Product Safety Authority (NVWA), which also conducts checks on products crossing the border, will become more similar and might eventually integrate or merge.

It is noted that in these configurations, the data formats are complex and inflexible. They are primarily designed at the European level and further specified at the national level. In this process, existing techniques and semantics are taken into account to enhance interoperability and create common terminology among countries. As respondent 3.22 of the case Single Window stated: "*The data quality was not so much an issue as definitions were set and knew no variation. (...) The market already knew the World Customs Organisation Data model!*" This ensures a broader understanding, which benefits the consistency and accuracy of the data set.

However, this implies that adapting formats at the international level requires significant effort. If a change is needed, it must be aligned at the European level and with the Organisation managing the data model. While the current system is functional, any future changes will require substantial time, expertise, and financial resources.

Secondly, regarding technical quality in the EU regulatory configuration, a certain degree of tension is observed between the different types of quality. As respondent 4.9 stated, "*The information infrastructure is very reliable. Because the system has been around for a while, most bugs have been found and fixed, making it very stable.*" However, this configuration faces challenges in modernising systems and components. Old legacy systems remain in use here.

It is important that the PII is reliable and accessible, as institutions are legally required to submit reports. As one can observe here, the preference is for proven technologies rather than progress. Some stakeholders in the PII favour change. However, change would require significant investment from both the government and the private sector, so it is often postponed and implemented at a moderate pace. Additionally, the receiving parties often develop and manage central systems. Due to the importance of reliability, accessibility, and uptime, control of these systems is usually kept in-house, according to the respondent of the Single Window case.

Finally, regarding governance quality in the EU-regulated configuration. According to the respondents in the EU-regulated configuration, governance is perceived as stable, largely due to the presence of a structured governance framework at the European level. Ownership is clearly divided within the PII, but

respondent 4.9 noted that the right knowledge is not always in the right place. An agency might own part of the PII, but this does not necessarily mean it understands the responsibilities that come with it. Decisions made at the European level may not match national infrastructures or the formats and standards in use. Therefore, it is essential that all stakeholders are actively involved in the decision-making process. Maintaining governance across multiple levels, both national and international, demands significant time and resources from the actors in the PII.

Although collaboration is regarded as necessary, only public organisations are often represented in the national governance structure, indicating little collective alignment with the private sector and limited inclusivity. Consequently, lobbying and bilateral meetings are crucial for fostering alignment within this context. One reason collaboration and agreements are less of a part of the governance might be the legal and regulatory foundation of the PII. The respondent of the Single Window case noted that “*A change almost only occurs when there is a legal basis.*” Focusing less on development through agreement and more on development through European directives and national regulations. This can increase technical efficiency, yet does not improve governance quality.

6.6.4 QUALITY IN THE MIXED COLLABORATION CONFIGURATION

In the mixed collaboration configuration, data quality is improved by using a standardised data format, which is viewed positively. This means the data is of technical quality and, because it is structured, easy to technically validate and process (R 4.2, 4.4, and 4.5). However, assessing the actual accuracy of the data within the PII remains challenging because no substantive validations occur. As the eInvoicing respondent said, “*It is like a mailbox that is open to spam as well. You do not know what ends up in the box.*” Respondents 4.4 and 4.1 observe this as an opportunity to implement assurance measures within the PII and consider this the next logical step for these configurations.

The numerous data formats provide flexibility for different stakeholders in the PII. However, some respondents mentioned that the variety of data formats and the necessary conversions can reduce data quality and incur additional costs. Furthermore, flexibility in a data format, such as using free fields, can decrease data quality but also allows for greater adaptability in the application. This approach enables the format to be adopted by parties that initially deemed it unfeasible (R4.6), thereby increasing its adoption and use.

Secondly, regarding technical quality in the mixed collaboration configuration, there is a high use of existing standards. Due to the high degree of standardisation, respondents 4.1, 4.3, and 4.5 stated that there is consistency within the configurations, and their scalability is high. Additionally, they consider the PII to be technically stable and mature, while recognising that these configurations are constantly evolving. Many have been redeveloped throughout their lifecycle. Although seen as mature, redeveloping the PII can lead to growing pains.

It is noted that the number of supported messages within the PII remains limited. The PII operates in a state between a chain and a network, as not all message types are supported, and there is limited diversity

in the supported standards (R4.5 and 4.1). While this presents opportunities for further growth, it still benefits from the stability, robustness, and accessibility of the existing proven systems and techniques (R4.2, 4.3, 4.4, 4.5, and 4.6). Over time, user satisfaction and ease of use decrease, increasing the urge to redevelop (R4.1 and 4.2). Changing parts of the PII can be a significant challenge due to the higher diversity of technical components. *“The larger the ecosystem, the harder it is to change. You have to win over parties to change with you”* (R3.3). The interdependencies between the systems are quite large. Additionally, this means issues and redevelopments are substantial alignment tasks and require good governance.

This means that a clear overview and comprehensive knowledge of the PII are crucial to understanding all interdependencies and improving specific components in harmony with the rest of the PII. This places considerable demands on the resources of the involved parties and can lead to failures when unexpected updates or releases occur in interconnected systems. Consequently, aspects such as chain logging and auditability become vital issues within the configuration (R4.5).

Finally, regarding governance quality in the mixed collaboration configuration, decision-making and development of PII do not take place in a central place. This makes it difficult to maintain high governance quality, as it demands significant coordination among the various bodies. Respondent 4.1 mentioned that it is necessary to have an overview of the different technical components, data formats, and all governance structures involved in the decision-making process. As respondent 4.5 stated: *“It is not always clear who is responsible.”*

Conversely, respondent 4.6 observed that transparent and relatively quick decision-making does occur within the individual governance structure. Only the directly involved stakeholders are represented, which speeds up the decision-making process. However, the challenge lies in whether stakeholders can effectively coordinate and align these decisions at a higher level. The process is often slow and cumbersome, as alignment is needed across multiple governance structures. As respondent five noted, *“If something goes wrong, it is searching for someone who can deal with it, and there is little decisiveness available to solve it”* (R4.1). This leads to negative perceptions of decision-making speed and can cause misalignment between different parts of the PII. The governance becomes more complex in a rapidly changing environment.

Compared to the earlier configuration, in mixed collaboration, cooperation occurs even without a formal legal basis. (R. 4.2, R 4.3). It is noteworthy, however, that the respondents feel that despite the complexity and lack of enforcement, parties know how to find each other and take responsibility. Yet, as respondent 4.3 stated: *“Even if you are of good will, mistakes are easily made and difficult to trace.”* Besides, cooperation cannot always be enforced. Compliance with the agreements and escalation, therefore, is challenging

The positive aspect of governance quality is its generally high level of inclusiveness, as it involves both public and private parties with diverse roles. However, this inclusiveness can also lead to conflicts when private interests compete with public ones. Without clear leadership or enforcement capabilities, resolving conflicts becomes more difficult.

6.6.5 QUALITY OF PUBLIC COLLABORATION CONFIGURATION

Regarding data quality in the public collaboration configuration, the respondents consider it reasonable. However, as in other configurations, only the technical validity of the data is verified, and there is no in-depth validation by intermediary organisations or service providers. The responsibility for data quality rests with the data providers, making enforcement challenging. Nonetheless, data quality is often seen as better when the providers have a vested interest in accuracy (R4.11). For new data exchange PII, it is usually assumed that data quality is low. Improvements only happen once stakeholders collaborate to discuss the data exchange content, which helps them better understand the data and its application (R4.11).

Secondly, regarding technical quality in the public collaboration configuration. In these configurations, there is a central system provided by the government. Respondents in the notification property value case observed that there was formerly a focus on government standards, but now there is a growing trend towards adopting market standards. This move ensures greater interoperability in the future and increases market knowledge of the standards, thereby reducing the risk of vendor-lock-in.

A central infrastructure provides the benefit of managing system quality collectively in a single location, rather than each organisation needing to maintain and further develop its own separate component (R3.2). However, fulfilling all customer requirements does increase the system's demands. While joint development can be efficient, this efficiency diminishes as complexity rises. In some cases, government-provided systems are not mandatory, and organisations can develop their own systems (R3.25). This offers greater flexibility within the playing field.

Finally, regarding governance quality, this configuration involves several different bodies in which public parties collaborate, although there are usually fewer than in the mixed collaboration configuration. Even with fewer bodies, alignment remains necessary between the various governance structures. This requires organisations to have resources connected to several governance structures to represent their interests, which complicates the decision-making process (R4.11). *"For work to be done, we depend on commitment, but commitment is based on voluntariness. So sometimes it is difficult to get something arranged. One party is always needed to take the lead."* (R4.11). Additionally, there is often a lack of a higher-level vision and steering for the PII (R3.25). Due to differing interests within the governance, decision-making can slow down, and prioritisation must be achieved through consensus.

As a risk, in this configuration, governance can become entangled in indecision. *"All options are explored, but no decisions are made. Everything is overthought, and by the time the government has made a choice, the technical world has already innovated and changed?"* (R3.25). This applies to the development of components of PII. Respondent 3.16, for example, mentioned that at the initial stages, it is often not well considered how new components will be incorporated into management and how governance will be organised in relation to existing administrations. In this regard, funding remains a major challenge (R3.16). However, other respondents observe that joint funding is more frequently available for further development and management within these configurations.

6.6.6 QUALITY IN THE DISTRIBUTED CONFIGURATION

Regarding data quality, the distributed PII configuration is perceived as having reasonable to good data quality. This is because data quality is a regularly discussed topic within the central agreement framework. Since the data format is developed centrally with input from all relevant parties (both public and private) and roles (users and technology partners), data quality is viewed as relatively high (R4.12). Furthermore, users of the PII benefit from enhanced data quality, such as through decreased administrative burden, which in turn encourages further efforts to improve data quality (R4.12). As respondent 4.13 observed, the central agreement framework is designed for ongoing, flexible discussions focused on continuous improvement. However, this process is slow, and data quality is likely to stay suboptimal until consensus is achieved. The Kik-v case study demonstrates that data is only shared when there is consensus within the consultation group that doing so is significant and worthwhile. No extra measures have been introduced to ensure data quality beyond mutual understanding and the technical data format. In this setup, data quality is not mandated but shaped by the alignment of stakeholder interests.

Secondly, regarding technical quality, within a distributed PII configuration, agreements and standards dominate the design of PII. All agreements are made centrally, ensuring data is exchanged consistently. This enables data providers to prepare effectively, knowing the process will be consistent across all participants (R3.15). It guarantees system quality on multiple levels. It makes the PII highly interoperable and robust (R4.12 and 4.13), as one system's failure does not affect the others. It also makes it easier for parties to develop and integrate exchange requirements into their own systems (R4.13). However, if central modules are still used, even temporarily, this can pose risks and increase dependency. The question remains whether this is the most efficient approach. System-level collaboration might potentially improve efficiency. Still, in practice, software is often supplied by a limited number of vendors who implement the technology for end users. This creates a risk of vendor lock-in (R4.13).

Finally, regarding governance quality, distributed PII configuration represents the most inclusive form of governance. However, it also involves many parties, each with its own interests (R4.12). A central agreement framework necessitates a certain level of stakeholder participation. When participation is low or representation is uneven, some parties may gain disproportionate influence. Essentially, these agreement-based systems aim to simplify complex many-to-many relationships into a manageable set of shared agreements. Therefore, it is vital that all parties adhere to these agreements to ensure balance and fairness (R3.15 and R4.12).

The distributed PII is not driven by regulated information exchange but by standardisation based on mutual agreements and trust. This makes it more difficult to meet key governance quality requirements such as sustainable funding, conflict resolution, efficiency, and decisive decision-making. The lack of enforceability can cause private parties to be hesitant to connect and implement. Anchoring in legislation or obligations could help address this (R4.12).

The governance can, however, be unclear and indecisive. Respondent 4.12 stated: *"In my view, there is far too much talk and too little action. Recommendations from the pilot are adopted only very slowly after very long talks and a*

review." As a result, current roles and competences are not clearly defined, making it difficult to ensure consistent progress in developing the PII (R4.13). This configuration calls for guidance to prevent role conflict between steering and lobbying parties (R4.12). For this, an independent steward could contribute to the quality of the PII.

6.7 CHAPTER CONCLUSION

This chapter presents the findings of research phase three *Configurations*, and identifies the data, technical and governance sub-configurations that collectively shape the prevailing configurations of PII. First, the analysis of data-related sub-configurations identified six distinct sub-configurations, ranging from entirely unstructured data exchange to fully structured data formats shaped by one or more stakeholders. Second, the examination of technical design choices uncovered four technical configurations, ranging from centralised infrastructures to distributed infrastructures. Third, governance sub-configurations at the information infrastructure level were analysed, revealing six governance sub-configurations that vary from single, centralised governance configurations to decentralised governance configurations involving multiple governance structures. In chapter 6.4, we zoomed in on single governance structures.

Next, we mapped the interconnectedness among the data, technical and governance sub-configurations. We then compare these sub-configurations across cases to identify patterns in PII configurations. Collectively, these sub-configurations form the basis for six prevailing PII configurations. Some of these are relatively recently designed and established, while others have been in place for a long time. To further assess these configurations, additional surveys were conducted to evaluate their respective strengths and weaknesses from practitioners' perspectives.

During this research, we have gained several insights into PII stewardship through analysing the authorities involved in designing, maintaining, and adapting the PII throughout its various life cycles, to ensure data, technical, and governance quality. To deepen our understanding of PII stewardship, the next chapter will provide a synthesis of previous findings.

Chapter 7

SYNTHESIS:
EMPIRICALLY DERIVED
CLASSIFICATION OF PII
STEWARDSHIP

In the previous chapters, we established the design variables relevant in PII configurations, PII sub-configurations, and prevailing PII configurations. PII stewards are responsible for shaping and ensuring the quality of these PII across data, technical, and governance dimensions throughout their lifecycle. This study adopted a qualitative, contemplative approach, merging a conceptual synthesis from research phase one *Definitions* with a secondary data analysis of the cases used in research phase three *Configurations*. Here, we integrate insights from the literature review, case studies, document analysis, survey results, and semi-structured interviews to give us a deeper understanding of the empirical-derived classification of PII stewardship. In this fourth research phase, *Stewardship*, we answer the sub-research question: **What classification of public information infrastructure stewardship can be distilled from the findings of this study?**

In Chapter Three, we discussed the concepts of PII and PII stewardship, drawing on relevant theoretical sources. Based on the twenty-two cases from research phase three and their configurations, we can analyse the stewardship roles linked to each configuration. This secondary data enables us to enrich our conceptual analysis of PII stewardship and to empirically derive to a classification of PII stewardship.

In this chapter, we provide insights into different configurations from a life-cycle perspective. As mentioned earlier in Chapter 3.1.5, information infrastructure development can be distinguished into multiple phases. Moreover, it is the task of the PII steward to maintain, adapt, and evolve the PII throughout its various life cycles. During our analysis of the cases, we identified the phases of the PII cases within the PII life cycle. The analysis revealed that some PII configurations are relatively new and face challenges that PII stewards must address, which PII in later life cycle phases do not, and vice versa. In Chapter 7.1, we discuss changes in PII configurations and how these can influence the choices PII stewards need to make. Based on this analysis, the concept of PII stewardship, and the analysis of Chapter Six, we identify that the classification consists of three distinct categories of PII stewardship. In Chapter 7.2, we provide an overview of these three categories and explore how they shape prevailing configurations and how the PII stewardship is shaped by the PII configurations.

7.1 CHANGES IN PII CONFIGURATIONS THROUGHOUT ITS LIFECYCLE

Given the complexity and dynamics of PII development, it is essential to invest in aligning the PII and ensuring its quality. Configurations must correspond with data exchange conditions, contextual factors, and ongoing developments. Over time, collaboration has led to more complex PII configurations, often involving multiple components managed by different actors. For many configurations, aligning the PII proves challenging, as decisions are not always made centrally and there is significant interdependence among components. Many of these fragmented configurations have already undergone multiple reshapes and evolved throughout the PII lifecycle.

In Figure 7-1, we plotted the configurations to the lifecycles in which they were identified. During this research, this distinction between the phases turned out to be applicable to the PII context. However,

in the case of eCodex, the respondent stated: “*In the worst-case scenario, another country could, of course, simply use e-codex as a post office box and still print it out.. (..) and it is not mandatory to use eCodex, yet?*” (R3.7). Without proper use or implementation of the PII according to its purpose, drift can also occur in the ‘expansion and momentum’ phase. Therefore, the configurational model is not yet suitable for the existing context, leading to misuse or underuse. We updated the model in Figure 7-1 to illustrate the possibility of drift occurring in both the ‘expansion and momentum’ and the ‘stagnation and inertia’ phase. In the following paragraphs, we will elaborate on the phases and the configurations identified during that phase.

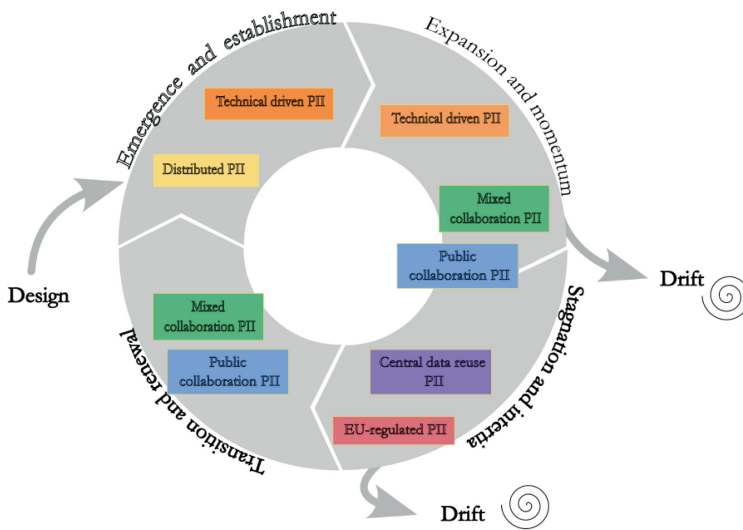


Figure 7-1. Configuration development through the life cycle, Adapted model on the life cycle, based on Bolton & Foxon, 2015; Edwards et al., 2009

Based on the analysis of the PII configuration, we can determine that the different cases are in various phases of the life cycle. The cases of eCodex, OOTS, KIK-v, and EDU-V provided insights into their newly established PII. In these cases, we observe the cases emerging and being established with a clear purpose. As stated by respondent 3.8: “*The aim is to establish a trust framework that standardises the education sector.*” Here, a new direction was taken to fully redevelop the way educational organisations exchanged information. In other cases, we observe the infrastructure being pushed towards adoption. As respondent 3.7 stated: “*It was still very little used. After a number of years, the infrastructure was complete, and attempts were made to encourage as many countries as possible to use it.*” However, existing literature on infrastructure lifecycle models (Bolton & Foxon, 2015; Edwards et al., 2009) indicates that this model includes four lifecycle phases: ‘emergence and establishment’, ‘expansion and momentum’, ‘stagnation and inertia’, and ‘transition and renewal’, which in turn evolve into ‘emergence and establishment’, and drift can occur during the ‘stagnation and inertia’ phase.

The public and mixed collaboration cases provided insight into their phases, as the respondents noticed that the PII was established, and some cases tend to reach the stagnation and inertia phase. As respondent

3.12 noted: “*There isn't really much more to say about LSP now, because it has been running stably for a number of years. There is not much change anymore*”. However, he continued that changes are still possible, “*This often concerns security and the measures that need to be taken to ensure it*.”(R3.12). The respondent from 3.2 noted: “*There are still areas for improvement. It is often pointed out that the quality needs to be improved even further. I think that will always be the case*.” The cases of public and mixed collaboration configurations, show they are flexible enough to develop the PII in the future, as the respondent 3.5 stated: “*I am actually focusing full-time on the further development of the system. We are trying to get the best out of it. We currently have an excellent programme for that*.”. The development, in turn, will be established in the PII and implemented by the stakeholders.

Some configurations are considered more difficult to redevelop. The cases of Central data reuse and the EU regulation of PII showed signs of difficulties during stagnation and inertia phases. As respondent 3.17 of the Wage Returns statistics case noted: “*They [the PII stewards] say, we have other priorities. You would actually want there to be sufficient capacity on an ongoing basis to move that information infrastructure forward*”. In these cases, there tends to be a lack of resources and priority to redevelop the PII. In both the Wage returns statistics and the Pension funds case, organisations with a strong interest in the case that lack formal stewardship roles or influence do not have decision-making rights and are unable to develop the PII themselves. As respondent 3.18 stated: “*We are dependent on data suppliers, so we need to be more flexible. We cannot simply say, 'I want it, and it must be done this way. ' We have to go with the flow.*”

We identify at least two trends in the development of the PII lifecycle that we wish to emphasise. First, it can be stated that cases designed in previous decades have predominantly focused on centralised infrastructures. This is especially evident in configurations categorised as Central data reuse or EU-regulated. These cases typically reside in the stagnation phase and do not appear in earlier phases of the infrastructure lifecycle. As stated in the literature, over time, technologies and processes may start to diverge from the original design objectives of the information infrastructure (Ciborra, 2004). The respondents have a general expectation that such infrastructures will remain stable until they become too unsustainable or outdated, rather than being proactively redeveloped or modernised. As respondent 3.17 stated: “*There is a possibility, however, it is rather pessimistic, that at some point they stop the data exchange (..) However, we must deliberately consider the reduction of burden, costs, and ease of use for civilians. We need each other*.” As stated earlier, this tendency is largely driven by a lack of prioritisation and insufficient resources dedicated to the further development of the infrastructure. As stated by respondent 3.19: “*At this point, we are pleased that data exchange occurs and we receive data. Standardisation has the lowest priority (..) and the physical exchange of information, how that works. That is also a really low priority*.” The risk is that, once the infrastructure fails, the costs of replacement will be significantly higher, likely requiring the development of a new PII. Delaying intervention may allow for the adoption of more advanced technologies at a later stage; however, postponement also increases the risk of drift, as well as higher implementation and stakeholder management costs. Whether to invest in maintenance or replacement should ideally be informed by a cost–benefit analysis to determine the most viable option.

Second, we are witnessing a shift in the design of newer information infrastructures toward decentralisation, or even full distribution of the PII systems. Particularly in the early stages of the life cycle, we observe configurations following this trend. These newer cases can be found in the distributed PII and

the technical-driven PII. As stated by respondent 3.23, *‘During the COVID period, we started collaboration with the private sector. We want to enable people to find jobs, and the idea is that private corporations benefit from this as well. (...) We therefore started a public-private collaboration’*. In this case, the configuration is formed by trust frameworks and a high level of collaboration. These configurations, including the cases of EDU-V, KIK-V, and VUM, are characterised by higher levels of market and end-user involvement, a strong emphasis on the use and reuse of standards, and a diminished focus on central systems. As stated by respondent 3.15: *‘We try to capture data exchange in agreements rather than in a central facility’*. This evolution significantly alters the governance landscape. The role of the government is no longer dominant; decision-making increasingly relies on broader stakeholder consensus. This often results in more inclusive, widely supported infrastructure. However, this also means that the government's control over the PII is reduced. Here, there is a challenge to the broader adoption of PII, and the prisoner's dilemma emerges. As respondent 3.8 stated: *‘Public and private collaborations are not necessarily self-evident. This is a very challenging way to establish the foundation based on complete equality.’* A critical mass of users is needed for the infrastructure to gain traction and establish agreements.

7.2 PII STEWARDSHIP

In Chapter Three, we gave the following definition for PII stewardship: *PII stewardship refers to the authorities for ensuring data, technical and governance quality in designing, maintaining, and adapting a public information infrastructure throughout its life cycles.*

According to the literature, information infrastructures are becoming increasingly larger and more complex (Hanseth & Monteiro, 1998). This is also evident in practice, for example, through collaboration, PII components can be widely used and reused by multiple organisations. We observed that technical components are utilised and reused by various parties, including the Digipoort, LSP, Single Window, SBR standards, i-standards, and the WCO data model. However, collaboration also carries risks. Components developed collaboratively often face higher requirements, and alignment between the needs is difficult as more organisations are involved. However, respondent 3.6 noticed: *‘What excites me is that we have a new governance structure, which is actually much clearer. Much clearer about what needs to be done. (...) This helps with the alignment of the actors’*. As collaboration grows, it is increasingly important to stay focused on ensuring all types of quality, and where there is often a focus on technical and data quality, this highlights the importance of governance quality.

Several researchers argue that information infrastructures cannot be fully controlled due to ongoing development and increasing complexity (Ciborra, 2004; Hanseth & Lyytinen, 2010). Here, we observe two elements of PII stewardship addressing this concern. First, the form and focus of PII stewardship are inherently dynamic, evolving as infrastructures change, new technologies emerge, and governance shifts. As theory shows, a PII is a constantly changing entity. Just as the configuration of a PII captures a moment in time, the stewardship of a PII is also subject to change, even though this can be a challenge for the authorities and actors involved. As respondent 3.22 stated: *‘I believe people need to recognise the importance and necessity of this change... Why are we making this change? Parties across Member States see that there is insufficient harmonisation of the data.*

In that sense, no one is opposed to change. However, with any significant adjustment, you often encounter a degree of fear or resistance.” Stewardship can evolve and adapt to reflect the evolving nature of the infrastructure. As a result, multiple forms of stewardship are possible, and the specific configuration may change as the infrastructure and its context evolve.

As defined in research by Otto (2011), *“While data owners are ‘accountable’ for the immediate correctness and consistency of certain data, data stewards develop and provide the rules for the handling of this data, i.e., they are ‘responsible’ for the overall data management?”* (p.8). From the cases, we observe that traditional PII stewardship authority is usually confined to a single organisation or part of an organisation. In cases such as the Pension Funds, Wage Return Statistics, Bank Supervision, Import, Migration Reporting, Single Window, and Land Registry, responsibility for PII is clearly assigned to a single dominant party primarily responsible for both the technical aspects and governance, as well as its quality. However, in most cases, PII stewardship reflects the distributed and interconnected nature of PII. There can be multiple authorities involved for maintaining quality, and stewardship is therefore shared among multiple (parts of) organisations in various ways. Figure 7-2 presents an overview of the classification consisting of three categories of PII stewardship.

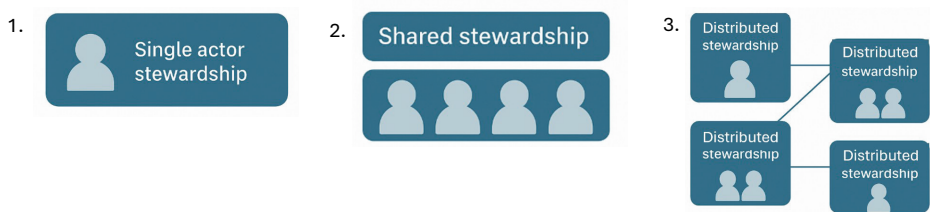


Figure 7-2. Overview of the three empirically derived categories of PII stewardship

In cases such as VUM, KIK-V, EDU-V, Carecodex, eCodex, and OOTS, there tends to be a single governance structure; however, in this governance structure, responsibilities for the PII are shared among different authorities. Respondent 3.15 stated on the central governance structure: *“Here, we work together and in that cooperation we adhere to the agreed agreement.”* In other cases, there is no single governance structure, but rather several governance structures, each with its own actors involved. Examples include Notification property value, eInvoicing, SBR-education, WMO, Vaccination, WOCO, BRP issue reporting, Child authority, National housing progress, and child authority. In these cases, we observe different sub-governances overseeing aspects such as the data format, technical components, and standards used in the PII. This highlights that stewardship can be fulfilled by multiple parties within a single governance structure or distributed across different parts of the PII.

Table 21. Overview of Stewardship types and related configuration and cases.

PII stewardship categories	Configuration	Corresponding cases
<i>Single actor stewardship</i>	Central data reuse PII	Pension funds, Wage returns statistics
	EU-regulated PII	Bank supervision, Import, Migration reporting, Single Window, Land registry
<i>Share stewardship</i>	Distributed PII	VUM, KIK-V, EDU-V, Carecodex
	Technical-driven PII.	eCodex, OOTS
<i>Distributed stewardship</i>	Mixed collaboration PII	Notification property value, eInvoicing, SBR-education, WMO, Vaccination, WOCO
	Public collaboration PII	BRP issue reporting, Child authority, National housing progress, and Child authority

Table 21 presents an overview of Stewardship categories and related configurations and cases, showing a relation between the stewardship types and the configurations in practice. In the following sections, we elaborate on these three forms of PII stewardship.

7.2.1 CATEGORY 1 - SINGLE ACTOR PII STEWARDSHIP

In single actor stewardship, as the name suggests, one organisation holds authority for the entire PII. PII stewardship refers to the authorities for ensuring data, technical and governance quality in designing, maintaining, and adapting a public information infrastructure throughout its life cycles. Here, the responsibilities are assigned to an individual or organisation for maintaining aspects such as security, integrity, usability, and description of data (Rosenbaum, 2010; Dawes, 1996). The stewardship role is central to ensuring effective quality management (Wende, 2007). In pension funds, wage return statistics, bank supervision, import and migration reporting, single window, and land registry cases, we observe a single authority filled by a single organisation. These cases represent central data reuse PII and EU-regulated PII cases. It can be observed that single-actor stewardship often relies on legislation and regulations, with the organisation responsible for their implementation performing this task fully. In central data reuse PII cases, law-making occurs at the national level, whereas in EU-regulated PII configuration cases, stewardship is mostly defined at the European level. The European level facilitates decision-making on technical standards, which must then be implemented at the national level. *"Ultimately, such legislation is eventually passed, and then it simply becomes a fact. And then it becomes a fait accompli here, which we simply implement. So the jobs and the business community consult, but customs is ultimately the one that decides. That is what the government prescribes."* (R.3.9). This makes it easier to implement and enforce decisions than with other categories of stewardship. The PII steward's tasks involve translating legislation into a PII of the highest possible quality. Through bilateral contacts or working groups, the PII steward can consult stakeholders and incorporate their views into the decision-making process. Respondent 3.9 continued: *"Between businesses and customs, is it more about informing and coordinating than making decisions together? I get the impression that the private sector does not have much say in this, but it would be useful if they were included."* Yet, the decision is made by a single actor. Although decision-making within this form of stewardship is generally quicker, it may take longer to embed decisions into legislation and regulations. Agreements are formalised and enforced through bilateral agreements and contracts. This

enables the PII steward to hold actors accountable for non-compliance with PII. For PII stewards, the primary focus is on ensuring data and technical quality. Their role centres on making decisions that uphold high standards and on ensuring the effective and qualitative implementation of relevant regulations. While coordination with other stakeholders is necessary, the emphasis remains firmly on maintaining the integrity and quality of data and technical systems.

This category of stewardship seems to suit simple PII with few actors. As the complexity and dependence on other systems, standards, and infrastructures grow, it can become increasingly difficult to organise single-actor stewardship. In single-actor stewardship, we do observe some shift towards making agreements instead of regulation. Respondent 3.19 noted: *“We have been working on a standard, so that everyone agrees on the same way of recording how data is exchanged. (...) Those kinds of standards are now in place.”* This may, over time, lead to more collaborative configurations and shared responsibilities; however, at this point, we still observe a strong PII steward. This is still driven by the initiative of individual actors. In addition to their regulations, coordination with public and private parties can be sought to ensure the PII’s effectiveness. Respondent 3.22 remarked: *“We have a MIG, where we have the message implementation guidelines for companies (...) Companies can also have a say in the design. But ultimately, customs make the proposal.”* This does require a great effort and resources from the PII stewards in organising and developing the PII.

Here, we observe that PII configurations driven by a single stewardship can lead to a period of stagnation and inertia without clear priorities. Respondent 3.19 noted: *“There will, of course, come a time when volumes can no longer be handled in the traditional way, and systems become obsolete. Technical updates to these systems are gradually becoming a real necessity.”* However, this respondent noted that the priority of the technical infrastructure is low. Without the PII steward allocating resources or having a sense of urgency for the PII, the configuration will remain stagnant, with a risk of drift. Here, urgency is often promoted by the actors in the PII through lobbying rather than collective decision-making.

7.2.2 CATEGORY 2 - SHARED PII STEWARDSHIP, WITH SUPPORTING ORGANISATION

In Chapter 3.2, we have established that PII stewardship can be shared among multiple authorities as the PII increases. From the cases, we identify the second observed category of stewardship: shared PII stewardship. In cases such as VUM, KIK-V, EDU-V, Carecodex, eCodex, and OOTS, there tends to be a single governance structure; however, within this governance structure, multiple authorities are involved in the shaping of the PII. In this form, stewardship is collectively held by a group of actors, often supported by a coordinating organisation or programme.

Data and technical quality are ensured by multiple authorities that reach mutual agreements. As these parties share a common interest in data, power is more balanced; instead of one dominant player, there is a collective understanding and shared commitment to developing and adhering to PII standards. *“The idea behind the collaboration is establishing a public-private partnership in the labour market. (...) There were several developments, but at one point we realised that we needed a mechanism to collaborate”* (R3.23). There is a common public value that aligns with the public sector’s objective, enabling easier collaboration.

Stewardship concerning technical and data quality is found within the governance structure or trust framework. For example, the trust framework in VUM defines the data format, ensuring data quality in the PII, sets standards for data exchange, and offers a central service for data transfer. Participants in the trust framework ensure technical and data quality, and each organisation implements the agreements within their own part of the PII. Agreements are made centrally but executed in a decentralised manner, enabling organisational flexibility.

In shared stewardship, there is an emphasis on what the parties can agree upon to improve the quality of the PII. *“In our case, agreements have been made between the information requesters themselves. They have collectively agreed on the requested information, and that this will be done in accordance with the agreements laid down in the trust framework.”* (R3.15). In this form of stewardship, the focus is on aligning stakeholders and collaboratively formulating agreements.

This can lead to longer decision-making processes, as Respondent 3.23 stated: *“Only the Catholic Church is slower and more sluggish, and let's be realistic. The Catholic Church needs a thousand years to get anything done. If we can do it in five years, then I think that's actually quite impressive”*. The length of the decision-making process is partly due to the variety of topics and the involvement of multiple stakeholders who wish to contribute their interests. However, once agreements are reached, they are usually broadly supported within the PII. This reduces the risk of misalignment and encourages quicker and more effective implementation, partly due to the active participation of stakeholders. Problems can also be identified and resolved more swiftly, thus enhancing the PII's responsiveness.

Two types of configurations relate to this category of stewardship. The first is the configuration of technical-drive PII's, including the cases of eCodex and OOTS. Here we observe that several parties, often at a European level, are involved in building an innovative common system. Respondent 3.21: *“A European commission has set up a team of experts on the topic of architecture. But also on the topic of security, governments, and all related aspects.”* The responsibility is shared among the organisations involved in the European governance structure, as this is where the PII develops. We expect, however, that as the PII's further evolve, the configuration will become more complex, involving more organisations during the expansion phase, which will quickly lead to a redeveloped PII.

The distributed PII configuration, however, explicitly chooses central shared stewardship as a future-proof model. In distributed stewardship, the design, maintenance, and ongoing development of the PII, as well as the quality assurance of its components, are centrally coordinated.

An advantage of centralised shared stewardship is the potential for self-regulation over time. Compliance with agreements is often supported by collaboration and collective decision-making, allowing stakeholders to hold one another accountable when standards are not met. However, enforcing agreements remains challenging due to the absence of formal laws and regulations. The case of Edu-V aims to address this by establishing the Edu-V certificate of approval. *“The quality mark actually shows that you, as a party, meet the requirements we have set, and that you comply with the agreements of the agreement system.”* (r3.8). End users can observe which providers hold the Edu-V certificate, which benefits them by ensuring interoperability, ease of use, and security. When governance is well-designed and actively maintained, decisions about data quality

and technical performance can be made more efficiently; however, without incentives or obligations to contribute, there is a risk that PII could be compromised.

In shared stewardship arrangements, we observe that governance quality often rests with a single authority. It is beneficial when this authority has no direct interest in using PII itself, but rather in ensuring its proper overall functioning. Respondent 3.8: *“We take an independent position. Because we are independent, you will never see us say ‘we believe from a public conviction that this must be done’. (...) This provides a balanced discussion (between public and private organisations).”* The independent supporting organisation has the task to encourage actors in the PII to contribute to the development of the PII and participate in the governance. Separating interests allows attention to be divided more effectively between data and technical quality, and governance quality. Respondent 3.15: *“We support the establishment of the trust framework and ensure the actors in the framework make agreements (on the technical component), and that they translate these agreements to their contracts with care providers”.* This means that the governance quality steward has to ensure that the PII is inclusive, with clear decision-making procedures and well-defined escalation mechanisms. Besides promoting the adoption of individual systems or standards, supporting organisations also play an active role in encouraging the adoption of the PII as a whole. This is especially important when the PII is in the establishment or expansion phase. The role of the supporting organisation in ensuring governance quality within the PII is crucial to its existence. However, it remains uncertain whether shared stewardship will continue to function effectively over time, as we have not yet observed any long-standing examples.

7.2.3 CATEGORY 3 - DISTRIBUTED PII STEWARDSHIP

During the analysis of the case, we establish that PII stewardship can be shared among multiple authorities, but at the same time can also be divided across different parts of the PII. This results in distributed PII stewardship. This was actually the largest group of cases, including Notification property value, eInvoicing, SBR-education, WMO, Vaccination, WOCO, BRP issue reporting, Child authority, National housing progress, and Child authority. In each of these cases, different parts of the PII have different PII stewards or multiple PII stewards. We observe this category of stewardship represented in the configurations of mixed collaboration and in public collaboration configurations. In the mixed collaboration PII, the stewardship is largely distributed. Multiple levels and governance structures exist whereby different parts of the PII are developed, established, and maintained. However, in the public collaboration PII, stewardship is far less fragmented. Here, collaboration has resulted in large PII components. These larger infrastructural components, although influencing the PII to a large extent, do not encompass the entire PII and do not have full shared decision-making authority for data exchange.

Although there is often a dominant party with a clear need for data, we observe in practice that stewardship becomes distributed across multiple actors through collaboration. In the case of eInvoicing, it becomes evident that a governance structure exists for creating a European data format with a semantic norm and accompanying syntax. OpenPeppol is responsible for parts of the PII infrastructure, specific data formats based on the European norm, and has its own governance. There is a central system on the public

side, Digipoort and ROAP, which enables a single entry point for receiving public organisations, each with its own governance. Additionally, there is public collaboration on the data format. All these PII components have their own PII stewards involved, who hold authority for that part of the PII. We observe that in these configurations, multiple systems, services, standards, data formats, and governance structures can coexist without a single coordinating authority across the entire PII. As a result, the authorities are distributed.

In cases such as WOCO, Vaccination, SBR-education, and eInvoicing, there is an intermediary organisation involved for managing standards or systems. These parts of the PII often have their own governance structure designed for evolving components that may be used beyond a specific PII. As the respondent of WOCO noted, aligning can be more of a challenge, as the intermediaries develop parts of the systems and standards but are not responsible for the overall process. Through collaboration, different actors can hold authority in the PII. For example, one or more parties may be involved in ensuring the quality or technical robustness of the PII. While control over parts of the PII can be enforced to some extent through laws and regulations, much also relies on mutual agreements and contracts. This makes enforcement more complex and emphasises the need for ongoing coordination and alignment of interests among the participating parties.

The study by Matthiesen and Bjørn (2015) shows that the legacy of government information infrastructure is often embedded within other frameworks, such as interconnected IT systems. This is especially difficult in these PII configurations. Therefore, distributed PII stewards need to have increased attention to the ongoing developments and the various dimensions of quality. *The point is that you want to be interoperable with as many parties as possible. And one network that offers this is the PEPPOL network. But it is not the only network available. (...) We do not know how this will develop.”*(R3.3) Due to the distributed nature of the PII, more effort is required to harmonise the various decentralised components. However, this attention often remains siloed, with each stakeholder focusing primarily on their respective part of the infrastructure.

All PII stewards need to monitor and ensure coherence among the components of the PII. A significant consequence of distributed stewardship is that decision-making occurs at various levels and among different stakeholders. Respondent 3.11 stated: *“Trying to establish national cooperation processes and cooperation protocols. Then, once that working process has been clarified, it also establishes the provision and exchange of information. (...) It is very difficult. Everyone has enough on their plate, each of those organisations.”* In such a complex field, seeking alignment and integration can be time-consuming and challenging. The speed of decision-making varies by level, depending on the number of stakeholders involved and the underlying legal frameworks.

7.2.4 RESPONSIBILITIES ACROSS STEWARDSHIP CATEGORIES

In this chapter, we aim to answer the research question: **What classification of public information infrastructure stewardship can be distilled from the findings of this study?** In the previous subsection, we established three distinct categories of stewardship: single-actor PII stewardship, shared PII stewardship, and distributed PII stewardship. In analysing these three categories of stewardship, we found

that each requires different knowledge. In this final sub-section, we compare the different categories of stewardship and provide insights into the knowledge needed by the PII stewards.

Table 22. Overview of responsibilities, knowledge and common issues for each stewardship category

PII stewardship categories	Involved Authorities	Knowledge	Common issues
<i>Single actor stewardship</i>	Central authority to ensure data, technical, and governance quality	High knowledge of both data, technical, and governance components	Priority and resources
<i>Shared stewardship</i>	Stewardship is shared among multiple authorities in a centralised structure.	Deep knowledge of Data and technical	Inclusiveness, having a representative governance.
	Often, a supporting authority is established to ensure the quality of governance.	Knowledge of governance management and quality.	
<i>Distributed stewardship</i>	Stewardship is shared among multiple authorities but divided across different parts of the PII.	Knowledge of how to ensure high quality of one's own part of the PII and achieve alignment with the rest of the PII.	Alignment between the different parts of the PII

In single-actor stewardship, a single organisation ensures the different types of quality. This requires greater effort and expertise within the organisation to organise and manage the PII. It requires a deep understanding of both the technical aspects of the PII used for data exchange, the content of the data, and its formatting, as well as the capabilities to align stakeholders' needs with the PII. We have seen cases where resources are too limited, it is a low priority, or it is too complex for a single organisation to take on the role of PII steward entirely. In these instances, we observed a primary focus on compliance through laws and regulations. When they steer the PII strictly within legal boundaries, this often leads to a reactive rather than proactive approach to stewardship.

In shared stewardship, stewardship is shared among multiple authorities in a centralised structure. Those involved often focus on data and technical quality in the PII. However, the drive for consensus and seemingly rational solutions can conceal significant differences in values, power, access to resources, and vulnerabilities among stakeholders, or in our case, stewards (Turnhout et al., 2020). This emphasises the need for stewards who are not only technically knowledgeable about the PII but also adept at navigating governance dynamics and managing the associated risks.

Alongside knowledgeable stewards, shared stewardship demands active participation from individuals. Without active participation, stewardship cannot truly be shared. In this context, proactive stewardship involves making agreements throughout the PII. Once these agreements are established, a lack of interest or resources from public parties can lead to decreased activity in the PII. One solution is to assign an

authority for governance quality, which helps keep the PII governance more engaged. This reduces the risk of stagnation. However, this also requires dedicated resources for an organisation that oversees the governance. Such a setup creates a division of expertise: the governance-enabling organisation can focus on establishing governance structures, while the actors involved concentrate on the content, including data and technical quality.

Finally, in distributed stewardship, stewards are required to possess capabilities to coordinate across organisations, with a clear vision for the future and practical expertise. They must balance the effort and resources they allocate between actively participating in other PII governance and developing the specific components within the part of the PII over which they have authority. After all, if every organisation continues to work in isolation, the risk of misalignment remains high. Once misaligned, it can demand more effort and resources to update existing processes to meet new standards.

This places high demands on PII stewards. They must have some knowledge of PII and be able to develop a strategic vision and facilitate coordination. This requires a different kind of PII steward, one who can operate effectively within a network of governance structures, stakeholders, and dependencies outside their part of the PII. In environments where centralised decision-making is less common, further development and innovation are more adaptable. However, this requires significantly more effort to manage the field effectively.

Generally, we can conclude that dedicated resources are necessary for the maintenance, management, and redevelopment of the PII. Configuration failures are mainly due to a lack of priority, capabilities, and resources for ongoing maintenance, management, and further development.

7.3 CHAPTER CONCLUSION

This chapter integrated the insights from the preceding chapters to analyse the complex environment in which PII stewards operate. The findings indicate that PII evolves, progressing through distinct phases in its life cycle. Some configurations are able to progress relatively easily through these phases, demonstrating stability and continuity. Others are still within the emergence and establishment phases, reflecting newer or still-developing configurations. Over time, these newer configurations have to prove their applicability, and only with enough adoption can the PII become established. In contrast, several configurations appear to struggle during the stagnation and inertia phases, where they risk drifting and becoming misaligned with their intended purpose.

The analysis in this chapter further demonstrates a clear relationship between PII configurations and categories of PII stewardship. Both the PII configuration is shaped by the PII stewards, and the configuration also influences the PII stewardship and the possible PII stewardship categories.

Three forms of PII stewardship were identified. The first, single-actor stewardship, is predominantly associated with the central data reuse PII and EU-regulated PII configurations. The second, distributed stewardship, is evident in the mixed collaboration and public collaboration PII configurations, where responsibilities and decision-making are shared across multiple organisations. Finally, shared PII

stewardship characterises the distributed PII and technical-driven PII configurations. These findings show that PII stewardship is not merely an organisational arrangement but is intrinsically linked to the configuration and evolution of PII itself. Different kinds of stewardship demand distinct capabilities from stewards. We can conclude that dedicated resources are necessary for the maintenance, management, and redevelopment of the PII.

Chapter 8

CONCLUSIONS AND RECOMMENDATIONS

In this study, we focused on data exchange with public organisations. In doing so, we examined the design variables of PII quality, PII configuration, the PII life cycle, and PII stewardship. In this chapter, we present our main findings and answer the main research question. This main research question is twofold: **What is PII stewardship, and how does it relate to configurations for data exchange with (semi-)public organisations?** Here, we discuss the scientific and practical contributions, outline the research limitations, reflect on the study, and offer recommendations for future research.

8.1 FINDINGS

This is the first research that investigates public information infrastructure configurations and the stewardship of public information infrastructures. To answer the main research question, we address six specific sub-research questions.

The first sub-research question aims to define and conceptualise PII, as it is an underexplored concept. The aim is to understand PII by identifying its key characteristics and how it differs from a PII, including the complexity involved in designing PII and its entire lifecycle. Therefore, the first sub-research question is: **What is a public information infrastructure?** We addressed this in Chapter 3, where we identified that public sector organisations need information that goes beyond their internal operations and business processes. Consequently, PII are established to retrieve or receive this information. PII are socio-technical systems consisting of interconnected technical and social components that facilitate information exchange within public organisations. They evolve over time rather than emerging spontaneously, being shaped by the decisions and interactions of the involved actors. They often rely on shared standards to ensure interoperability and stability.

Due to the increased complexity and configurational dynamics of a PII, steering is necessary to ensure alignment and focus on quality. This type of steering should focus on the design and redesign of PII, as well as the impact of design choices on the final infrastructure and PII quality. In Chapter 3, we addressed the second sub-research question: **How can public information infrastructure stewardship be defined?** We established that the complexity of PII is extremely high. The challenges for governments include managing increased dependence on external private organisations, conflicting values between organisations, multi-stakeholder decision-making, and heightened technical complexity. To address these challenges and given the complexity, it is essential to have authorities for maintaining, adapting and evolving the PII throughout its life-cycles. Drawing on concepts of governance and data stewardship, we propose a new form of authority for PII: *Public Information Infrastructure Stewardship*. PII stewardship offers a new approach that bridges the gap between operational data stewardship and higher-level governance. PII stewardship involves the authorities for guiding and maintaining the quality of a PII in terms of data, technical, and governance quality. This stewardship is exercised throughout the entire lifecycle of the PII, covering the design, implementation, expansion, and transition phases.

Based on the literature, we establish that there are various types of stewardship possible. Due to the involvement of multiple actors in the PII, collaboration can be pursued among different authorities, whereas other PII can enable single-actor stewardship.

Furthermore, different stewards may define varying quality objectives. To evaluate the quality of a PII, it is essential to understand what constitutes PII quality. Drawing on various quality categories in PII stewardship, we developed a quality framework that includes the dimensions and measures of PII quality. We identified three key dimensions: data quality, technical quality, and governance quality. Each dimension comprises several measures to assess the quality of the PII.

Building on these ideas, we assert that PII stewards possess the authority to shape the PII design. These choices represent specific implementations of design variables. A design variable defines a dimension along which alternatives can be selected. We have considered the design choices that impact the overall design of a PII, including the information exchanged, technical components such as systems, standards, and services, and the governance structures involved. Additionally, we focused on the responsibilities of these elements, as they influence the distribution of responsibilities and, thereby, PII stewardship. To gain a clearer understanding of the concept of PII and the range of decisions that PII stewards can make, we needed to address sub-research question three: **What are the relevant design variables of a public information infrastructure?** (Chapter 5).

Through a literature review and a case study with interviews and document research, we established a list of twenty-nine design variables divided into three categories: Data, technical, and governance. Our analysis identified ten relevant data design variables. The first variable concerns whether a formatted data structure exists, and if so, whether multiple formats are permitted. The next variable considers the complexity of these formats, regarding semantics, syntax development, technical support, or a combination of these factors. The creation of these formats may be done individually or collaboratively and can be tailored for data exchange or based on a standardised data model. We observe that reusing formats is common. The final variable addresses who is responsible for creating and managing the data formats.

For the technical component of the PII, we identified eleven relevant design variables, covering systems, standards, and services. The first variable relates to the relationship between senders and receivers. In some cases, a 'report once, used by many' approach is feasible. The PII can involve multiple or single providers and recipients, which can provide or retrieve data from its source. The relevant systems may have central, semi-central, decentralised, or distributed architectures. Collaboration can occur at both the system and standards levels. During data exchange, one or multiple services may be used upon request by the actors involved in the PII. Finally, both public and private actors can supply the technical components within the PII.

We have identified nine governance design variables through our analysis. First, consider the number of governance structures and the extent of these structures. According to the literature and interview data, in public governance configuration, multiple structures are common, often forming hierarchical, networked, or combined ordering. These governance structures involve a variety of actors, often including both public and private entities, such as suppliers, receivers, intermediaries, regulators, policymakers, and third parties.

Enforcement of decisions within these governance structures varies and is frequently linked to the funding of the PII.

These relevant variables demonstrate that PII configuration designs must be developed carefully and emphasise the importance of considering design variables collectively, rather than in isolation. We expect each case to have made different design choices, leading to diverse configurations. However, we anticipate that some choices will be more common in certain instances, forming prevailing configurations. Using the established framework of relevant design variables, we examined the configurations of these design choices in practice in Chapter 6 and addressed sub-research question four: **What are the prevailing configurations of PII through which organisations exchange data with (semi-)public organisations?**

Based on our analysis, we identified that there are data sub-configurations, technical sub-configurations, and governance sub-configurations. Based on these sub-configurations, we found six prevailing configurations. These PII configurations include ‘Central Data Reuse’, ‘EU-Regulation’, ‘Technical Driven’, ‘Public Collaboration’, ‘Mixed Collaboration’, and ‘Distributed with Trust Framework’.

The first configuration, the central data reuse PII, is characterised by a centrally organised system that involves minimal cooperation between involved parties. A key feature of this configuration is that the initiation of data exchange is established at the national level by regulation. Receiving organisations require data to fulfil their statutory duties, and this data is obtained directly from central government agencies. The coordination between the parties involved in this configuration is primarily bilateral and limited in scope.

The second configuration, the EU-regulated PII, is also characterised by a centrally organised system with limited cooperation among relevant actors. However, here the initiation of data exchange occurs not at the national level, but at the European level, usually through EU regulations or directives. In this context, the Netherlands has established so-called shadow committees to oversee the national implementation and impact of these European frameworks. The lead in designing the PII typically lies with either the sending or receiving party.

Configuration three, the technical-driven PII, originates from European initiatives that are primarily technical in nature. These initiatives focus on establishing technical exchange mechanisms for cross-border data exchange, and harmonise exchange structures, such as transport protocols, authentication and addressing. In practice, this often involves unstructured document exchange.

Configuration four, the public collaboration PII, is characterised by centralised cooperation in the field of digital systems, where government-wide standards play a key role. This configuration focuses on developing and applying public standards, which are widely (re)used within the public sector. In recent years, there has been a clear shift towards greater market involvement and an increasing focus on market standards.

Configuration five, the mixed collaboration PII, is characterised by a hybrid information infrastructure that combines both centralised systems and decentralised or distributed components. This configuration shows a high level of complexity, with multiple links in the chain involving both public and private intermediaries. The infrastructure supports a broad range of services, including, for example, audit reports with hash functions, technical checks on data formats, identification systems, white and blacklist services for authorisation, and data transformations.

Configuration six, the distributed PII, is characterised by its distributed nature and the strong emphasis on trust frameworks as a structure for coordination. In this configuration, the number of central systems is kept to a minimum. Instead, focus is on developing and managing common standards for data exchange and formats through explicit governance structures. Decision-making is based on consensus, with jointly established agreements forming the foundation for interoperability within the infrastructure.

In Chapter Three, we elaborated that PII quality consists of three dimensions: data quality, technical quality, and governance quality. Each configuration will have different outcomes, strengths, and weaknesses of these quality types. In the second part of Chapter Six, we explored these differences and answered sub-research question five: **What are the strengths and weaknesses of the prevailing configurations of public information infrastructures in terms of their quality?**

We recognised that different configurations have varying strengths and weaknesses. The configuration of central data reuse focuses on data reuse. As a result, the data-supplying party tends to be less aware of the importance of data exchange, even though it is this party that has developed the systems for sharing. Consequently, the prioritisation of quality within the chain remains relatively low. However, the techniques are proven and implemented robustly, and all participants are familiar with the standards and systems, leading to minimal disruption. Only through new initiatives and renewed attention can the quality improve over time.

In the technical-driven configuration, there is a strong emphasis on technical quality, while data and governance quality receive comparatively little attention. Little attention is paid to data formats, yet technical solutions are adopted to improve data quality. The governance quality on the European level is deemed high, yet the national governance is still in development. The focus on technical quality resulted in a technically robust system; however, the system's utilisation remains a point of uncertainty.

In an EU-regulated configuration, data quality is typically viewed positively. Existing data models, validation services, and international alignment contribute to this positive view. The technical quality is also viewed positively, as it uses proven methodologies and standards. The infrastructure is stable. However, updating or fixing existing systems, formats, and standards tends to be difficult. They are inflexible and require discussion at both the national and international levels, necessitating significant alignment. Although the governance is clear, inclusiveness is low. Lobbying and bilateral meetings play a significant role in this alignment.

In the mixed collaboration configuration, there are generally positive results regarding the quality of the PII. Due to the many data formats and technical components, flexibility in use is key. Additionally, this configuration is deemed mature and capable of evolving through the life cycle. However, because of legacy systems and the variety of standards, data formats, and technical components, development is challenging. Interdependencies are high, and reaching agreement among the different stakeholders and governance structures takes time. Without dedicated resources, misalignment poses a greater risk.

In the public collaboration configuration, there is extensive collaboration on data formats, which improves data quality. Additionally, the involvement and interest of the sending parties in the data exchange process further enhance this quality. A large emphasis is placed on understanding the data to support quality

improvement. Due to the central systems and alignment with market standards, the system quality generally remains high. Shared services and systems can improve efficiency and facilitate the use of higher-quality technologies through collaboration. However, a trade-off exists between meeting requirements, managing complexity, and maintaining efficiency. Although multiple governance structures still exist, they are fewer than in the mixed collaboration configuration. Nonetheless, alignment stays important, and strategic high-level guidance is recommended to steer the overall vision.

And finally, in the distributed configuration, ensuring the quality of PII primarily focuses on aligning the interests of stakeholders. There is a central agreement framework that includes many stakeholders with different roles, ensuring that all decisions are aligned. However, aligning interests and resolving conflicts becomes more challenging as there is no enforcement. This means that, rather than emphasising technical and data quality—which result from stakeholder interests—more effort is needed to ensure governance quality and protect public values.

In the previous chapters, we identified the design variables in PII configurations, sub-configurations, and prevailing PII configurations. PII stewards are tasked with shaping and ensuring the data, technical, and governance quality of these PII throughout its lifecycle. By synthesising the theoretical conceptualisation of PII stewardship and the data collected in the third research phase regarding PII configurations, we can answer the final sub-research question: **What classification of public information infrastructure stewardship can be distilled from the findings of this study?**

By analysing the twenty-two cases, we identified three types of stewardship. First, in single actor stewardship, as the name suggests, a single central authority shapes the PII and ensures data, technical, and governance quality. We observe that single actor stewardship primarily relies on legislation and regulations, and that the organisation responsible for implementing them carries out this task entirely. This makes it easier to implement and enforce decisions compared to other types of stewardship. The PII steward's role involves translating legislation into high-quality PII. The second type of stewardship is distributed PII stewardship. Although there is often a dominant party with a clear need for data, we observe in practice that stewardship becomes distributed across multiple authorities. What may start as a regulatory requirement can develop into a more cooperative approach, with parties actively seeking alignment. This can offer benefits such as increased effectiveness or a simplified exchange process for providers. Stewardship is shared among multiple authorities, but at the same time, divided across different parts of the PII. While control over parts of the PII can be enforced to some extent through laws and regulations, much also relies on mutual agreements. This complexity underscores the need for continuous coordination and alignment of interests among the involved parties.

The third and final type of stewardship observed is shared PII stewardship. Stewardship is shared among multiple authorities in a centralised structure, often supported by a coordinating organisation or programme structure. Data and technical quality are ensured by multiple authorities that reach mutual agreements. As these parties share a common interest in data, power is more balanced; instead of one dominant player, there is a collective understanding and shared commitment to developing and adhering to PII standards.

Answering all sub-research questions provides us insight into PII, its configuration and the stewardship needed. The main objective of this research was to answer the main research question: **What is PII stewardship, and how does PII stewardship relate to configurations for data exchange with (semi-) public organisations?** Since the main question is two-fold, we answer this question separately.

First, the question of **What is PII stewardship?** Based on a conceptual exploration in theory and an empirical establishment of PII types, PII stewardship can be defined as: *‘the roles associated with the responsibility for maintaining, adapting and evolving the information infrastructure throughout its various life cycles, with a focus on ensuring data, technical and governance quality.’* There can be multiple types of PII stewardship. In our research, we identified three types of PII stewardship in practice: single-actor stewardship, shared stewardship, and distributed stewardship. In these different stewardship types, the involved authorities, required knowledge for the PII stewards, and challenges vary.

The second part of the main question is: **How does PII stewardship relate to configurations for data exchange with (semi-) public organisations?** We addressed this question by analysing how PII stewardship interacts with configurations throughout the PII’s lifecycle. We observe that each type of stewardship aligns with specific PII configurations. These configurations evolve over time, and as they do, the type of PII stewardship also shifts.

First of all, single-actor PII stewardship often results in configurations such as centrally reused data PII and EU-regulated PII. These cases typically reside in the stagnation phase and do not appear in earlier phases of the infrastructure lifecycle. These configurations seem to have been designed in previous decades. There is a general expectation that such infrastructures will stay stable until they become too unsustainable or outdated, rather than being proactively redeveloped or modernised. In these cases, we observe parties lobbying to influence the PII steward to develop the PII. Only with enough urgency and momentum do these cases change.

Secondly, distributed PII stewardship tends to lead to either public collaboration or mixed collaboration configurations. These configurations can be found in all phases of the lifecycle. We notice that configurations using distributed stewardship tend to be quite resilient throughout their lifecycle. In many cases, these infrastructures have already gone through several redesigns or updates. Our results show that, because of the distribution of the authorities, better alignment is needed among the different parts of the PII. A distributed PII steward must keep a broad, cross-domain view. The participation of multiple actors, each capable of initiating development, is a key force driving the evolution of the PII. While achieving alignment requires considerable effort, the resulting configurations seem to be robust and sustainable.

Finally, shared PII stewardship often leads to decentralised PII configurations. The cases involving distributed PII are relatively new and are typically observed only in the initial phase of the lifecycle. In shared stewardship, an independent organisation can provide and guide the governance, ensuring the governance quality of the PII. Stakeholders of the PII can participate in the governance structures and are encouraged to contribute to the development of the PII. The focus is therefore on what the actors themselves can agree on to improve the quality of the PII. These configurations are characterised by higher levels of market and end-user involvement, a strong emphasis on the use and reuse of standards, and a reduced focus on

centralised systems. The role of the government becomes less dominant; instead, decision-making tends to depend on broader consensus among stakeholders. This often leads to more inclusive and widely supported infrastructures.

Based on the results, it can be concluded that cases developed in previous decades predominantly focused on centralised infrastructures, particularly those falling under the categories of central data reuse and EU-regulated configurations. These infrastructures were generally expected to remain stable until they became unsustainable. Over time, many of these systems have evolved into more collaborative forms, including mixed and public-sector PII due to collaboration. These configurations have proven to be both sustainable and robust, despite the considerable effort required from PII stewards for alignment.

Currently, we observe a shift in the design of newer information infrastructures towards greater decentralisation, and in some cases, full distribution. In this context, the government's role is no longer dominant; instead, decision-making depends on broader consensus among stakeholders. How these distributed models will develop in the long term remains uncertain, but they mark a shift from conventional centralised approaches.

8.2 CONTRIBUTIONS

8.2.1 SCIENTIFIC CONTRIBUTION

This research started by expanding existing literature on government and information sharing from e-government (e.g. Layne & Lee, 2001; Linders, 2012), open government data (e.g. Dawes et al., 2016; Janssen et al., 2012; Purwanto et al., 2020) or the use of (big) data analytics within the government (e.g. Janssen et al., 2017; Kim et al., 2014; Klievink et al., 2017) with conceptualisation of public information infrastructure, PII design choices and PII configurations. We thereby broadened the focus from a single organisation (Gil-Garcia & Sayogo, 2016) to a multi-collaborative perspective that explores the public side of the information infrastructure literature. This research provided insight into the development of an updated PII life cycle model. This dissertation contributes to the academic understanding of public infrastructure governance by introducing a typology of six infrastructural configurations and providing the first overview of PII sub-configurations and prevailing PII configurations. It thereby establishes a foundational understanding of the patterns in which these configurations appear. By identifying these design choices that form PII configurations for the first time, this research enables further studies to develop testable hypotheses and provides a basis for future comparison. Moreover, highlighting the alignment problem faced by PII stewards in PII configuration adds a new layer to the existing literature by framing alignment not as a design-time challenge but as an ongoing institutional and technical effort that persists across the lifecycle of digital infrastructures.

Additionally, this research focuses on a new concept of public information infrastructure stewardship. Over the past decades, government agencies have become large data organisations (Janssen & Kuk, 2016; Klievink et al., 2017). The government finds itself in a complex position: while it has a public

responsibility to deliver high-quality services, it relies heavily on data flows originating outside its organisational boundaries. Drawing upon existing literature and complexities in practice, we have identified the necessity for a clear and comprehensive definition of PII stewardship.

Neither the concept provided in literature on governance (e.g. Stoker, 1998; Khatri & Brown, 2010; Klievink & Janssen, 2014; Weill & Ross, 2005; Wende, 2007), nor the concept of data stewardship (e.g. Dawes, 1996; Dawes, 2010; Rosenbaum, 2010; Wendelborn, 2023; Otto, 20011), was deemed sufficient to explain this phenomenon. The definition of PII stewardship is different from both concepts, as it focuses on the authorities for designing, maintaining, and adapting PII throughout its various life cycles, for ensuring data, technical and governance quality. This research, in particular, examines PII stewardship by linking it to PII quality and the PII life-cycle.

Moreover, the study also offers a concrete operationalisation of infrastructural quality through the dimensions of data quality, technical quality, and governance quality. In this research, we established that no existing model exists for PII quality. Based on literature on Data stewardship (e.g. Dawes, 2010; Rosenbaum, 2010; Wende, 2007), and Governance (e.g. Khatri & Brown, 2010; Klievink & Janssen, 2014; Weill & Ross, 2005; Wende, 2007), we established the three dimensions of PII quality as data quality, technical quality and governance quality. Based on established literature measures of quality—including works by Delone and McLean (1992), Fisher and Kingma (2001), Fricke and Schulz (2005), Gorla et al. (2010), Khatri and Brown (2010), and Lockwood (2009)—we developed a new comprehensive conceptual model for PII quality. For each of these dimensions, we explored the relevant quality measures, resulting in a list of twenty-five quality measures. This framework enables future research to examine PII quality and provides a structure for developing new hypotheses.

8.2.2 SOCIAL CONTRIBUTION AND IMPLICATIONS FOR PRACTITIONERS

For practitioners, this dissertation provides actionable insights into the long-term stewardship of public information infrastructures. There are three key takeaways for practitioners we want to underline.

First of all, the study emphasises that PII-stewardship should not be treated as a secondary or temporary function, but as a structurally embedded role that requires dedicated resources, authority, and technical expertise. Organisations involved in cross-organisational or even cross-country data exchange are encouraged to identify or appoint stewards to monitor alignment across the components of the PII. Alignment refers to the continuous need to coordinate and recalibrate the relationship between infrastructure, governance, quality, and lifecycle management. As this study demonstrates, alignment is not a linear process in which either the infrastructure is developed and governance subsequently aligned, or vice versa. Rather, it is a dynamic, co-evolutionary process, a continuous interplay without a singular solution. Based on the empirical analysis, four core alignment challenges can be identified. First, the need to configure and stabilise infrastructural arrangements across the entire PII. Second, the institutionalisation of PII-stewardship. Third, continuous monitoring and ensuring of PII quality (encompassing data quality, technical

quality, and governance quality), and fourth, managing the infrastructure lifecycle. Crucially, the alignment problem is not a one-off issue to be ‘solved,’ but a permanent challenge of re-alignment: *It remains ever-present.*

Next, practitioners should be aware of investing in adaptive governance and PII structures, so they are capable of evolving in response to technological developments and changing stakeholder landscapes. Infrastructures operate in rapidly evolving socio-technical environments, requiring them to be both technologically adaptable and governance-resilient. Governance configuration and the governance structures need to be arranged to be flexible and capable of responding to change, and at the same time, the infrastructure itself must be built to adapt to changing technologies, standards, and stakeholder constellations. Effective PII-stewardship plays a pivotal role in this, ensuring that continuous resource allocation, quality assurance, and strategic development are not treated as peripheral tasks, but as core elements of infrastructural maintenance.

Finally, more effort should be put into selecting knowledgeable stewards and training such stewards. Knowledge development and training are crucial, as professionals responsible for PII management must be equipped to handle complex, distributed environments and the ongoing demands for alignment. The changing form of stewardship requires a shift in professional and organisational capabilities. There is a growing need for knowledge management practices that support the development of expertise in managing complex, evolving infrastructures. Practitioners must be equipped with the capabilities and insights needed to navigate the interplay between infrastructure, governance, and their role as stewards.

It is essential that, in practice, resources are allocated for the management and maintenance of PII. Without clear priorities and dedicated attention, PIIs are likely to become adrift or evolve in ways that may not align with the objectives of governmental organisations. This principle applies to all forms of stewardship.

8.3 REFLECTION, LIMITATIONS AND FURTHER RESEARCH

8.3.1 REFLECTION ON THE UNIT OF OBSERVATION

This research provides a novel overview of public information infrastructures. However, five years ago, we initially started our research by examining public-private data ecosystems. It turned out this definition was multiple-interpretable, with limited boundaries and no specific orientation towards data sharing. The term was too broad; therefore, we specified the unit of observation to public-private information infrastructures. After the first two research phases, we found that not every information infrastructure that provides data sharing to public organisations involves private organisations. For Phase Three, we extended the concept to public information infrastructures.

However, our scope includes PIIs that support data sharing with (semi-)public receiving organisations. We took the perspective of governments in the cases. As one of the criteria, this research examined only cases in which the information is required for the primary processes of the information receiver, as there is a significant dependence on the information in these instances. The focus on governments as data recipients

is merely a subset of PII, as public organisations can also be data suppliers to private organisations. However, we exclude government-to-business (G2B) and government-to-citizen (G2C). The shift from public-private data ecosystems to public-private information infrastructure to public information could introduce a flaw in the research; however, we ensured that all cases used in the research represented PII. Therefore, we can extend the results towards the concept of PII.

8.3.2 REFLECTION ON CONFIGURATIONAL ANALYSIS

In this research, we aimed to examine the configurations of PII. Configurations are abstractions of reality that help us understand the possibilities designers have when designing a PII. By simplifying complex realities into variables and configurations, they help structure our understanding of the world and can provide insights into the underlying mechanisms shaping organisational arrangements.

Donaldson (1996) calls configurations “simplistic caricature” (p.127). He noted that organisations in the real world rarely fit into these configurations and are often in the middle between them or a combination of configurations. In our research, each configuration has its limitations, as models are reflections and simplified representations of reality. Nevertheless, conceptual models and configurations can still serve as useful analytical tools. Here we see a tension between complexity versus the reductionist view. A model should be as simple as possible so that many people can understand it. However, when simplified too much, the model does not reflect reality anymore. If there are many variables, the model might better reflect reality; however, it will be too complicated to understand. If there are too few variables in the model, it does not correspond to reality. As scholars, we must continuously search for the right level of abstraction.

In the research, the final number of clusters was determined based on a combination of dendrogram analysis, a visual inspection of the cluster matrix, and a qualitative assessment of clusters for relevance. During the analysis, we found that there is some overlap between these configurations and that the number of identifiable configurations depends on the level of granularity applied. However, based on the visualisation and qualitative analysis, we define the identified (sub)configurations as distinct configurations, sharing certain common design choices. On a higher level, several configurations can be categorised due to common design choices, giving a higher-level perspective.

Many configurational researchers view organisations as tightly coupled configurations that rarely change. If these configurations change, it will be a radical, fundamental, dramatic change (Demers, 2007). However, note that configurations are less tightly coupled and can slowly evolve. We expect that some configuration found, such as the technical-driven PII, over time, will evolve into other configurations, depending on the design choices made. This can imply that more configurations can appear over time, as technology advances and new governance structures can be formed. Based on our research we state that some of the changes in the configuration can be made easily, whereas other changes will be drastic and difficult.

Additionally, there are more possible configurations, or hybrid configurations. For example, we have already seen that eInvoicing and Single Window are configurations that can be combined or fall in between. Unlike strict configurational analysis, we observe that, due to their fluid characteristics, PII have more fluid configurations. However, there is a significant common denominator among the distinct configurations that we have identified in our research. We have also observed that certain configurations have existed for a considerable period, and have often evolved into alternative configurations over time, and new types of configurations are now emerging. It is expected that configurations will continue to evolve over time due to, for example, new technological developments, new preferred values influencing policy-making, changes in law and regulation, and European developments. Consequently, longitudinal research into the various configurations of PII would be a promising avenue for further research.

8.3.3 REFLECTION ON CONTEXTUAL FACTORS AND CRITICAL REALISM

This study adopts a critical realist perspective, which assumes that reality exists independently of the researcher, although the way phenomena are expressed can differ across contexts. Within critical realism, causal explanations refer to the mechanisms that produce outcomes under specific contextual conditions. The configurations identified in this thesis, however, should mainly be understood as descriptive rather than fully explanatory.

At this stage of the research, it is not possible to establish a causal effect of context through clear conditional statements. The study remains primarily focused on describing observed patterns and identifying relevant factors. The configurations identified reflect common combinations of design variables across the data, rather than comprehensive causal mechanisms that explain outcomes in various contexts. From a critical realist perspective, such descriptive and exploratory work can be viewed as a necessary step towards uncovering underlying mechanisms. The configurations presented in this thesis illustrate the relationships between design variables across data, technical, and governance domains, which collectively influence different forms of PII arrangements.

Context remains a crucial element in designing and developing PII. Besides these elements embedded in design choices, several overarching contextual factors are identified as either consistent across cases or different among them. In the studied cases, such factors influence various aspects of design decisions. For example, reusing standards depends on their prior existence, collaboration requires coordination and agreements among involved organisations, and adopting specific standards relies on the broader ecosystem in which organisations function. Future research can examine how contextual factors influence the development and governance of PIIs.

8.3.4 REFLECTION ON THE CLASSIFICATION OF PII STEWARDSHIP

In this study, we established the term "PII stewardship" and defined an empirically derived classification. We started with the definition of PII stewardship as the authorities for maintaining, adapting, and evolving the Public information infrastructure throughout its various life cycles, to ensure data, technical, and

governance quality. In doing so, we took data stewardship as our starting point. This involves the authority transferred to a specific actor representing (a part of) an organisation. Therefore, we started with assigned stewardship, in which authorities are mandated or established by rules and regulations. During this research, however, it turned out that, in the absence of proactive decision-making, actors in the PII can (try to) engage in PII stewardship roles out of a sense of urgency.

We mostly observe this single-actor stewardship. Once implemented, priorities and resources might be lacking, leading to stagnation and possible drift. Actors influenced by the PII will phase in and engage in stewardship out of a sense of urgency. This can eventually result in shared or distributed stewardship, thereby outdating single-actor stewardship.

Distributed stewardship, on the other hand, provides space for stewardship from urgency. Actors in the field who are influenced by decision-making can participate in governance structures, thereby influencing consensus-based decision-making.

Then the danger is that only parties who feel a sense of urgency can participate in decision-making, and parties without a sense of urgency become subordinate to these decisions, only later experiencing the consequences of not actively participating in the decision-making process. Therefore, it is essential that resources are allocated and PII stewardship is taken seriously.

8.4 LIMITATIONS

In this research, we encountered three limitations that we will discuss in this section.

The first limitation is that we cannot make claims about causal relationships in this research. The approach used in this research is critical realism. This approach holds that reality exists but is complexly layered and only partially understandable. The research focuses on identifying combinations of conditions rather than loose causal relationships. Although we observed correlations between PII configuration and PII design choices, this research makes no claims about causal relationships. We clarify how PII configurations and PII stewardship relate, and how patterns occur within configurations. While some initial indications of potential relationships have been noted, they have not been tested or validated. Future research should explore the interrelationship between design choices in more detail and, through longitudinal studies, investigate whether certain design choices or configurations lead to others over time.

A second limitation arises from the research approach. We examined 22 cases within the Netherlands. Through these cases, we aimed to explore a wide range to uncover underlying truths. However, these findings remain specific to the Dutch context. This study does not include a cross-country comparison; research was conducted solely within the Netherlands. While we identified some European-installed PII and European-driven PII, these are still highly contextual to Europeans, and thus, applicability to other countries may be limited. To some extent, we increased generalisability by using multiple cases across diverse domains, including youth care, healthcare, trade, public reporting, internal processes, and the labour market. Additionally, in cases such as OOTS, e-Invoicing, and Migration, collaboration is already

sought internationally. Additional research is needed to assess the broader applicability of the PII configurations and PII stewardship in other countries and to other cases.

The third limitation is that we only examine PII with a focus on (semi-) public organisations receiving data. This means the categories of stewardship identified are specifically related to public information infrastructure stewardship with a focus on the governments as recipients. While the definition of PII stewardship adds to the existing literature, it may be too narrow an interpretation of the concept. In this study, we primarily focus on publicly driven infrastructure in which public entities play a motivating role and where broader public interest in data exchange is evident. We suggest that research in G2X data sharing can add to the conceptualisation of PII. Furthermore, we propose that future research could include a private sector perspective alongside insights from the public sector, as this study focused on the government's viewpoint due to its dependence on information for essential processes and the need for tools to manage a multi-actor environment.

8.5 FUTURE RESEARCH

For future research, we identify three possible directions. First, we encourage further research on (P)II across various domains, including replicating the study and validating or complementing the given design choices and configurations. In this research, the variables are identified through three cases: eInvoicing, SBR, and SW. We validated the design choices in twenty-two other cases; although we found no evidence of missing variables, other cases or future cases might involve different or additional design choices. Our research offers a preliminary framework for these choices and configurations, which may evolve. The list of design variables and configurations is not fixed; new choices may emerge, and new configurations may be discovered in the future. Moreover, as new variables or design choices appear, some variables may become obsolete, leading to the development of new configurations and the discontinuation of others. Therefore, we recommend either repeating this study or establishing a longitudinal one. This would enable us to learn more about the further development of PII. It would also be interesting to observe whether a recognisable pendulum swing appears over time, given that history demonstrates that centralisation and decentralisation of systems constantly alternate. When one approach fails or faces too many issues, the opposite is often adopted. It would be valuable to investigate whether these shifts are becoming more moderate over time or if they continue to recur structurally. Additionally, the question remains which design variables influence these shifts. Currently, for example, we are witnessing a significant move towards collaboration, decentralisation, and modularity, partly enabled by open standards. Future research will uncover what the next phase or 'next best thing' might be.

Secondly, further research is necessary to assess the quality and success of PII. During the third phase of the study, we observed the quality associated with different configurations. However, this type of research proved to be more complex than anticipated, prompting us to adopt a more straightforward approach by examining the strengths and weaknesses in terms of data quality, technical quality, and governance quality. Through this research, we gained valuable insights into the strengths and weaknesses of PII configurations.

Extensive research focusing solely on quality is needed to validate the PII quality framework or even to extend it into a PII success framework. This research should consider the various interests and roles of public and private actors, such as intermediaries, recipients, senders, regulators, and policymakers. It is crucial to explore their specific interests and how they define and measure the success of a PII. With this knowledge, PII stewards can guide efforts towards quality outcomes that better align with the expectations of different stakeholders. If certain quality goals are unattainable, stewards can explain to participants why some interests are given lower priority at that moment from a perspective of public stewardship.

Finally, this research provided a first conceptualisation of PII stewardship. During the theoretical conceptualisation, we established that there will be different PII stewardship categories depending on the collaboration in the PII. Based on this assumption, we identified three categories of PII stewardship. Based on the insights from the research, we expect that information infrastructures not focused on data sharing with public organisations but towards, for example, private organisations, can have different PII stewardship categories. More research is needed on the different categories of stewardship, beyond the perspective of the public sector. This includes exploring how various stewardship models operate in different contexts, their impact on the provided quality and overall value, and the specific roles of stakeholders involved.

In sum, this manuscript shows that the development of public information infrastructures requires more than the design of temporary technical solutions or one-off institutional arrangements. They demand continuous alignment efforts and dedicated PII stewardship to ensure that infrastructures evolve with their context, maintain their quality, and remain fit for public purpose. While there is no universal blueprint, the prevailing configurations and categories of PII stewardship offered in this manuscript provide a foundation for both scholars and practitioners to better understand and shape these information infrastructures in an increasingly digital society

References

- Aanestad, M., Grisot, M., Hanseth, O., & Vassilakopoulou, P. (2017). Information infrastructures and the challenge of the installed base. *Information infrastructures within European health care: Working with the installed base*, 25-33.
- Aanestad, M., & Jensen, T. B. (2011). Building nation-wide information infrastructures in healthcare through modular implementation strategies. *The Journal of Strategic Information Systems*, 20(2), 161-176.
- Aben, T. A., Van Der Valk, W., Roehrich, J. K., & Selviaridis, K. (2021). Managing information asymmetry in public-private relationships undergoing a digital transformation: the role of contractual and relational governance. *International Journal of Operations & Production Management*, 41(7), 1145-1191.
- Aggarwal, C.C., & Reddy, C.K. (Eds.). (2014). *Data Clustering: Algorithms and Applications* (1st ed.). Chapman and Hall/CRC.
- AlShamsi, M., Salloum, S. A., Alshurideh, M., & Abdallah, S. (2020). Artificial intelligence and blockchain for transparency in governance. In *Artificial intelligence for sustainable development: Theory, practice and future applications* (pp. 219-230). Cham: Springer International Publishing.
- Babbie, E. R. (2020). *The practice of social research*. Wadsworth Cengage.
- Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)*, 41(3), 1-52.
- Baxter, P., & Jack, S. (2008). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. In *The Qualitative Report* (Vol. 13).
- Bharosa, N. (2022). The rise of GovTech: Trojan horse or blessing in disguise? A research agenda. *Government Information Quarterly*, 39(3), 101692.
- Bharosa, N., Hietbrink, F., Mosterd, L., & Van Oosterhout, R. (2018). Steering the adoption of Standard Business Reporting for cross-domain information exchange. In *Proceedings of the 19th annual international conference on digital government research: Governance in the data age* (pp. 1-10).
- Bharosa, N., Janssen, M., Hulstijn, J., van Wijk, R., de Winne, N., & Tan, Y. H. (2011, September). Towards a lean-government using new IT-architectures for compliance monitoring. In *Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance* (pp. 147-156).
- Bharosa, N., Janssen, M., van Wijk, R., de Winne, N., Van der Voort, H., Hulstijn, J., & Tan, Y. H. (2013). Tapping into existing information flows: The transformation to compliance by design in business-to-government information exchange. *Government Information Quarterly*, 30, S9-S18.

Bharosa, N., Janssen, M., Klievink, B., & Tan, Y. H. (2013, June). Developing multi-sided platforms for public-private information sharing: design observations from two case studies. In *Proceedings of the 14th Annual International Conference on Digital Government Research* (pp. 146-155).

Bharosa, N., van Wijk, R., de Winne, N., & Janssen, M.F.W.H.A. (Eds.) (2015). *Challenging the chain. Governing the automated exchange and processing of business information*. IOS Press.

Bhaskar, R. (1975). *A Realist Theory of Science*. Leeds Books.

Brooks, M. R. (2004). The governance structure of ports. *Review of network economics*, 3(2).

Bolton, R., & Foxon, T. J. (2015). Infrastructure transformation as a socio-technical process—Implications for the governance of energy distribution networks in the UK. *Technological forecasting and social change*, 90, 538-550.

Bostrom, R. P., & Heinen, J. S. (1977). MIS problems and failures: A socio-technical perspective. Part II: The Application of Socio-Technical Theory, *MIS Quarterly*, 1(4), 11

Bovaird, T., & Löffler, E. (2003). Evaluating the quality of public governance: indicators, models and methodologies. *International review of administrative sciences*, 69(3), 313-328.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In *International handbook of internet research* (pp. 97-117). Dordrecht: Springer Netherlands.

Braa, J., Hanseth, O., Heywood, A., Mohammed, W., & Shaw, V. (2007). Developing health information systems in developing countries: the flexible standards strategy. *Mis Quarterly*, 31(2), 381-402.

Brinkerhoff, D. W., & Brinkerhoff, J. M. (2011). Public-private partnerships: Perspectives on purposes, publicness, and good governance. *Public administration and development*, 31(1), 2-14.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done?. *Qualitative research*, 6(1), 97-113

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14-26.

Chen, H. (2002). Special issue digital government: technologies and practices, 34, 223-227.

Chen, T., Yu, Y., & Duan, Z. (2019, October). Data access & sharing approach for trade documentations based on blockchain technology. In *2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)* (pp. 1732-1736). IEEE.

Cheong, L. K., & Chang, V. (2007). The need for data governance: a case study. *ACIS 2007 proceedings*, 100.

Cherns, A. (1976). The principles of sociotechnical design. *Human relations*, 29(8), 783-792.

Ciborra, C. (2004). *Digital technologies and the duality of risk* (No. DP 27). ESRC Centre for Analysis of Risk and Regulation.

Cordella, A. (2010). Information Infrastructure: an actor-Network Perspective. *International Journal of Actor-Network Theory and Technological Innovation*, 2(1), 27-53.

Dawes, S. S. (1996). Interagency information sharing: Expected benefits, manageable risks. *Journal of policy analysis and management*, 15(3), 377-394.

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government information quarterly*, 27(4), 377-383

Dawes, S. S., Vidiyasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs : An ecosystem approach. *Government Information Quarterly*, 1–13.

de Bruijn, H., & ten Heuvelhof, E. (2008). *Management in Networks: On Multi-actor Decision Making*. Routledge.

de Graaf, G., & Paanakker, H. (2015). Good Governance: Performance Values and Procedural Values in Conflict. *American Review of Public Administration*, 45(6), 635–652.

De Winne, N., Janssen, M., Bharosa, N., Van Wijck, R., & Hulstijn, J. (2011). Transforming public-private networks: An XBRL-based infrastructure for transforming business-to-government information exchange. *International Journal of Electronic Government Research*, 7(4), 35–45.

Dearnley, C. (2005). A reflection on the use of semi-structured interviews. *Nurse Researcher*, 13(1), 19–28

Delone, W. H., & McLean, E. R. (1992). The quest for the dependent variable. *Information Systems Research*. *Information System Research*, 3(1), 60–95.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9–30.

Demers, C. (2007). The Configurational Approach. *Organizational change theories: A synthesis*. sage. 47–60.

Devaney, L. (2016). Good governance? Perceptions of accountability, transparency and effectiveness in Irish food risk governance. *Food Policy*, 62, 1-10.

Dutch Government. (2020, April 21). *NL DIGITAAL 2020 – Actualisatie Data Agenda Overheid* (Ministry of the Interior and Kingdom Relations). Retrieved July 31, 2021, from <https://www.digitaleoverheid.nl/document/nl-digitaal-2020-actualisatie-data-agenda-overheid/>

Dutch government (2025) Mededingswet. *Staatsblad* Last retrieved 19 October 2025 from <https://wetten.overheid.nl/jci1.3:c:BWBR0008691&z=2025-09-01&g=2025-09-01>

Easton, G. (2010). Critical realism in case study research. *Industrial marketing management*, 39(1), 118-128.

Edwards, P. N., Jackson, S. J., Bowker, G. C., & Knobel, C. P. (2007). Report of a workshop on history & theory of infrastructure: lessons for new scientific cyberinfrastructures. *Underst. Infrastruct. Dyn. Tens.*

Edwards, P. N., Bowker, G. C., Jackson, S. J., & Williams, R. (2009). Introduction: an agenda for infrastructure studies. *Journal of the association for information systems*, 10(5), 6.

Enjolras, B. (2009). A governance-structure approach to voluntary organizations. *Nonprofit and voluntary sector quarterly*, 38(5), 761-783

Ernst, A. (2019). How participation influences the perception of fairness, efficiency and effectiveness in environmental governance: An empirical analysis. *Journal of Environmental Management*, 238, 368-381.

European Parliament & Council of the European Union. (2016, April 27). *Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General*

Data Protection Regulation). Official Journal of the European Union, L 119, 1-88. Retrieved from <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>

European Parliament & Council of the European Union. (2018). *Regulation (EU) 2018/1724 establishing a single digital gateway to provide access to information, procedures and assistance and problem-solving services*. Official Journal of the European Union.

European Parliament & Council of the European Union. (2014). *Regulation (EU) No 910/2014 on electronic identification and trust services for electronic transactions in the internal market*. Official Journal of the European Union.

European Parliament & Council of the European Union. (2022a). *Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act)*. Official Journal of the European Union.

European Parliament & Council of the European Union. (2022b). *Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector (Digital Markets Act)*. Official Journal of the European Union.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. Wiley.

Fang, Z. (2002). E-government in digital era: concept, practice, and development. *International journal of the Computer, the Internet and management*, 10(2), 1-22.

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs - Principles and practices. *Health Services Research*, 48(6 PART2), 2134–2156.

Filer, T. (2019, January). *Thinking about GovTech: A brief guide for policymakers*. Bennett Institute for Public Policy, University of Cambridge

Fisher, C. W., & Kingma, B. R. (2001). Criticality of data quality as exemplified in two disasters. *Information and Management*, 39(2), 109–116.

Fricke, E., & Schulz, A. P. (2005). Design for changeability (DfC): Principles to enable changes in systems throughout their entire lifecycle. *Systems Engineering*, 8(4).

Fu, Y., & Zhu, J. (2019). Operation mechanism for g2b system based on blockchain. *Tehnicki Vjesnik*, 26(6), 1841–1852

Furukawa, M., & Minami, A. (2013). A Study on the 'Flexibility' of Information Systems (Part 1): Why Do They Need to Be Flexible? *International Journal of Business and Management*, 8(20), 48–61.

Gebauer, J., & Schober, F. (2006). Information System Flexibility and the Cost Efficiency of Business Processes 1. *In Journal of the Association for Information Systems* (Vol. 7, Issue 3).

Stoker, G. (1998). Governance as theory: five propositions. *International social science journal*, 50(155), 17-28.

Gil-García, J. R., & Pardo, T. A. (2005). E-government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly*, 22(2), 187–216.

Gil-García, J. R., & Sayogo, D. S. (2016). Government inter-organizational information sharing initiatives: Understanding the main determinants of success. *Government Information Quarterly*, 33(3), 572–582.

Gorla, N., Somers, T. M., & Wong, B. (2010). Organizational impact of system quality, information quality, and service quality. *Journal of Strategic Information Systems*, 19(3), 207–228.

Hanseth, O. (2010). From systems and tools to networks and infrastructures—from design to cultivation: Towards a design theory of information infrastructures. In *Industrial Informatics Design, Use and Innovation: Perspectives and Services* (pp. 122–156). IGI Global.

Hanseth, O., & Lyytinen, K. (2010). Design theory for dynamic complexity in information infrastructures: The case of building internet. *Journal of Information Technology*, 25(1), 1–19.

Hanseth, O., & Monteiro, E. (1998). Understanding information infrastructure. *Unpublished manuscript*, 27.

Haug, A., Zachariassen, F., & van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, 4(2), 168–193.

He, H., Zheng, Y., & Yang, Y. (2012). Research and design of heterogeneous data exchange system in E-Government based on XML. In *Advances in Computer Science and Information Engineering: Volume 2* (pp. 361–366). Berlin, Heidelberg: Springer Berlin Heidelberg.

Henfridsson, O., & Bygstad, B. (2013). The Generative Mechanisms of Digital Infrastructure Evolution1. *MIS quarterly*, 37(3), 907–931.

Henningsson, S., Gal, U., Bjørn-Andersen, N., & Tan, Y. H. (2011). The next generation information infrastructure for international trade. *Journal of Theoretical and Applied Electronic Commerce Research*, 6(1), 1–15.

Hofman, W. (2019). Toward large-scale logistics interoperability based on an analysis of available open standards. In *Proceedings of the I-EISA Conferences* (Vol. 9, pp. 249–261). Springer International Publishing.

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, 27(4), 371–376.

Janssen, M. (2025). Responsible governance of generative AI: conceptualizing GenAI as complex adaptive systems. *Policy and Society*, 44(1), 38–51.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management (ISM)*, 29(4), 258–268.

Janssen, M., Chun, S. A., & Gil-Garcia, J. R. (2009). Building the next generation of digital government infrastructures. *Government Information Quarterly*, 26, 233–237.

Janssen, M., & Estevez, E. (2013). Lean government and platform-based governance—Doing more with less. *Government Information Quarterly*, 30, S1–S8

Janssen, M., & Kuk, G. (2016). Big and open linked data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, 26(1–2), 3–13.

Janssen, M., Voort, H. Van Der, & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338–345

Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: an introduction to cluster analysis. John Wiley & Sons.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152.

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, 57(3), 78–85.

Klievink, A.J., Bharosa, N., & Tan, Y. (2016). The collaborative realization of public values and business goals: Governance and infrastructure of public-private information platforms. *Government Information Quarterly: an international journal of information technology management, policies, and practices*, 33(1), 67-79

Klievink, B., & Janssen, M. (2014). Developing Multi-Layer Information Infrastructures: Advancing Social Innovation through Public-Private Governance. *Information Systems Management*, 31(3), 240–249.

Klievink, B., Janssen, M., & Tan, Y. H. (2012). A stakeholder analysis of business-to-government information sharing: The governance of a public-private platform. *International Journal of Electronic Government Research*, 8(4), 54–64.

Klievink, B., Janssen, M., & Tan, Y. H. (2012, June). Blurring public-private boundaries: governance of information sharing in global trade networks. In *Proceedings of the 13th Annual International Conference on Digital Government Research* (pp. 183-191).

Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers*, 19(2), 267–283.

Krauss, S. E., Hamzah, A., Omar, Z., Suandi, T., Ismail, I. A., Zahari, M. Z., & Nor, Z. M. (2009). Preliminary investigation and interview guide development for studying how Malaysian farmers' form their mental models of farming. *The Qualitative Report*, 14(2), 245.

Kurnia, R. A., Praditya, D., & Janssen, M. (2019, May). A comparative study of business-to-government information sharing arrangements for tax reporting. In *International Working Conference on Transfer and Diffusion of IT* (pp. 154-169). Cham: Springer International Publishing.

Layne, K., & Lee, J. (2001). Developing fully functional E-government: A four stage model. *Government Information Quarterly*, 18(2), 122–136

Lee, T. Y., Yee, P. K., & Cheung, D. W. (2009, April). E-government data interoperability framework in Hong Kong. In *2009 International Conference on Interoperability for Enterprise Software and Applications China* (pp. 239-244). IEEE.

Le Grand, J. (1990). Equity versus efficiency: the elusive trade-off. *Ethics*, 100(3), 554-568.

Lewin, R. (1999). *The soul at work : unleashing the power of complexity science for business es*. Orion Business Books.

Linder, S. H. (1999). Coming to terms with the Public -Private Partnership. *American Behavioral Scientist*, 43(01), 35–51.

Linders, D. (2012). From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly*, 29(4), 446–454.

Luciano, J. S., Sayogo, D., Ran, W., DePaula, N., Jarman, H., Tayi, G., ... & Luna-Reyes, L. F. (2017). Building a certification and inspection data infrastructure to promote transparent markets. *International Journal of Electronic Government Research (IJEGR)*, 13(4), 53-75.

Lockwood, M. (2009). Good governance for terrestrial protected areas: A framework, principles and performance outcomes. *Journal of Environmental Management*, 91, 754–766.

Matthiesen, S., & Bjørn, P. (2015). Why replacing legacy systems is so hard in global software development: An information infrastructure perspective. *CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, 876–890.

Mermoud, A., Keupp, M. M., & Percia David, D. (2019). Governance models preferences for security information sharing: an institutional economics perspective for critical infrastructure protection. In *International Conference on Critical Information Infrastructures Security* (pp. 179-190). Cham: Springer International Publishing.

Meyer, A. D., Tsui, A. S., & Hinings, C. R. (1993). Configurational Approaches to Organizational Analysis. *Academy of Management Journal*, 36(6), 1175–1195

Miller, D., & Friesen, P. H. (1977). Strategy-making in context: ten empirical. *Journal of Management Studies*, 14(3), 253–280.

Monteiro, E., & Hanseth, O. (1996). Social Shaping of Information Infrastructure: On Being Specific about the Technology. In *Information Technology and Changes in Organizational Work* (pp. 325–343).

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40(2), 120–123

Närman, P., Schönherr, M., Johnson, P., Ekstedt, M., & Chenine, M. (2008). Using Enterprise Architecture models for system quality analysis. *Proceedings - 12th IEEE International Enterprise Distributed Object Computing Conference, EDOC 2008*, 14–23.

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, 21(4), 199–235.

OECD. (2025). *Government at a Glance 2025*. OECD Publishing. <https://doi.org/10.1787/0efd0bcd-en>

Ofe, H., & de Reuver, M. (2024). Rethinking Openness in Data Platforms: The Impact of Data Artifact Characteristics on Platform Openness: Consequences, Scope and Mechanisms. *Business & Information Systems Engineering*, 1-12.

Organisation for Economic Co-operation and Development. (2002). *Models of public budgeting and accounting reform: Volume 2 Supplement 1*. OECD Publishing.

Ottens, M., Franssen, M., Kroes, P., & Van De Poel, I. (2006). Modelling infrastructures as socio-technical systems. *International Journal of Critical Infrastructure*, 2(2), 133–145.

Otto, B. (2011). A morphology of the organisation of data governance. *19th European Conference on Information Systems*, ECIS 2011, March

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. In *The BMJ* (Vol. 372). BMJ Publishing Group.

Paide, K., Pappel, I., Draheim, D., & Vainsalu, H. (2018). On the systematic exploitation of the Estonian data exchange layer X-road for strengthening public-private partnerships. *ACM International Conference Proceeding Series*, 34–41.

Pawson, R., & Tilley, N. (1997). An Introduction to Scientific Realist Evaluation. In *Evaluation for the 21st Century: A Handbook* (pp. 405–418). SAGE Publications, Inc.

Peterson, R. R. (2004). Integration strategies and tactics for information technology governance. In *Strategies for information technology governance* (pp. 37–80). IGI Global Scientific Publishing.

Pipek, V., & Wulf, V. (2009). Infrastructuring: Toward an integrated perspective on the design and use of information technology. *Journal of the Association for Information Systems*, 10(5), 1.

Plugge, A., Nikou, S., & Janssen, M. (2022). A fuzzy-set qualitative comparative analysis of factors influencing successful shared service center implementation. *Industrial management & data systems*, 122(4), 920–941.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1), 109–130.

Provan, K. G., & Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory*, 18(2), 229–252.

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020). Citizens' trust in open government data: A Quantitative Study about the Effects of Data Quality, System Quality and Service Quality. *ACM International Conference Proceeding Series*, 310–318.

Robinson, N. J., Dahlhaus, P. G., Wong, M., MacLeod, A., Jones, D., & Nicholson, C. (2019). Testing the public–private soil data and information sharing model for sustainable soil management outcomes. *Soil use and management*, 35(1), 94–104.

Rosenbaum, S. (2010). Data governance and stewardship: designing data stewardship entities and advancing data access. *Health services research*, 45(5p2), 1442–1455

Rowley, J. (2002). Using Case Studies in Research. *Management Research News*, 25(1), 16–27.

Rukanova, B., Tan, Y. H., Huiden, R., Ravulakollu, A., Grainger, A., & Heijmann, F. (2020). A framework for voluntary business-government information sharing. *Government Information Quarterly*, 37(4).

Rukanova, B., Ubacht, J., Van Engelenburg, S., Tan, Y. H., Geurts, M., Sies, M., Molenhuis, M., & Slegt, M. (2021). Realizing value from voluntary business-government information sharing through blockchain-enabled infrastructures: The case of importing tires to the Netherlands using TradeLens. *ACM International Conference Proceeding Series*, 505–514.

Ryazanova, O., Pétercsák, R., Heaphy, L., Connolly, N., & Donnellan, B. (2016). Perception of Value in Public-Private Ecosystems: Transforming Dublin Docklands through Smart Technologies. Working paper for Thirty Seventh International Conference on Information Systems, Dublin 2016.

Sandra van Thiel. (2013). Bestuurskundig onderzoek. *Uitgeverij Coutinho*.

Scannapieco, M., & Catarci, T. (2002). Data Quality under the Computer Science perspective. *Computer Engineering*, 2(2), 1–12.

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2023). Artificial intelligence governance for businesses. *Information Systems Management*, 40(3), 229-249.

Scitovsky, Rudolf, Sabo, Kristian., Martínez-Alvarez, Francisco., & Ungar, S. (2021). Cluster analysis and applications . Springer.

Selsky, J. W., & Parker, B. (2005). Cross-sector partnerships to address social issues: Challenges to theory and practice. *Journal of management*, 31(6), 849-873.

Shin, D. H., Kim, W. Y., & Lee, D. H. (2006). Future public information infrastructure: lessons from four US case studies. *info*, 8(3), 47-59.

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, 7(1), 111–134.

Susha, I., D. Rukanova, B., Ramon Gil-Garcia, J., Hua Tan, Y., & Gasco, M. (2019, June). Identifying mechanisms for achieving voluntary data sharing in cross-sector partnerships for public good. In *Proceedings of the 20th Annual International Conference on Digital Government Research* (pp. 227-236).

Tan, P.-Ning., Steinbach, Michael., & Kumar, Vipin. (2006). Introduction to data mining. Pearson Addison Wesley.

Thompson, Grahame. (2003). Between hierarchies and markets : the logic and limits of network forms of organization. *Oxford University Press*.

Tilson, D., Lyytinen, K., & Sorensen, C. (2010, January). Desperately seeking the infrastructure in IS research: Conceptualization of "digital convergence" as co-evolution of social and technical infrastructures. In *2010 43rd Hawaii International Conference on System Sciences* (pp. 1-10). IEEE.

Tsang, E. W., & Kwan, K. M. (1999). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management review*, 24(4), 759-780.

Turnhout, E., Metze, T., Wyborn, C., Klenk, N., & Louder, E. (2020). The politics of co-production: participation, power, and transformation. *Current opinion in environmental sustainability*, 42, 15-21

United Nations. (2018). United Nations E-government survey 2018: Gearing E-government to support transformation towards sustainable and resilience societies. <https://www.preventionweb.net/publications/view/59447>

van de Kaa, G., Janssen, M., & Rezaei, J. (2018). Standards battles for business-to-government data exchange: Identifying success factors for standard dominance using the Best Worst Method. *Technological Forecasting and Social Change*, 137, 182–189.

van de Kaa, G., Van den Ende, J., De Vries, H. J., & Van Heck, E. (2011). Factors for winning interface format battles: A review and synthesis of the literature. *Technological Forecasting and Social Change*, 78(8), 1397–1411.

van der Peet, L., van Donge, W., Bharosa, N., & Janssen, M. (2025, August). Designing for Trust in Healthcare Data Sharing: Trust Anchors in the Trust Framework Lifecycle. In *International Conference on Electronic Government* (pp. 150-163). Cham: Springer Nature Switzerland.

van Donge, W., Bharosa, N., & Janssen, M. F. W. H. A. (2020). Future government data strategies: Data-driven enterprise or data steward?: Exploring definitions and challenges for the government as data enterprise. *ACM International Conference Proceeding Series*, 196–204.

Van Donge, W., Bharosa, N., & Janssen, M. F. W. H. A. (2022). Data-driven government: Cross-case comparison of data stewardship in data ecosystems. *Government Information Quarterly*, 39(2).

Van Donge, W., Janssen, M. F. W. H. A., & Bharosa, N. (2024, August). Preparing Public Agencies for Harnessing AI: A Study on Variables Shaping Multi Actor Information Infrastructures. In *International Conference on Electronic Government* (pp. 254-269).

van Engelenburg, S., Janssen, M., & Klievink, B. (2015). Design of a business-to-government information sharing architecture using business rules. In *SEFM 2015 Collocated Workshops* (pp. 124-138). Berlin, Heidelberg: Springer Berlin Heidelberg.

van Engelenburg, S., Janssen, M., Klievink, B., Tan, Y. H., & Rukanova, B. (2018, May). Comparing the openness of archetypical business-to-government information sharing architectures: balancing advantages of openness with the control of risks. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age* (pp. 1-10).

van Loenen, B., Kulk, S., & Ploeger, H. (2016). Data protection legislation: A very hungry caterpillar. The case of mapping data in the European Union. *Government Information Quarterly*, 33(2), 338–345.

Vest, J. R., Campion, T. R., Kern, L. M., & Kaushal, R. (2014). Public and private sector roles in health information technology policy: Insights from the implementation and operation of exchange efforts in the United States. *Health Policy and Technology*, 3(2), 149–156.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy : What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–33.

Watson, B., Fico, F., Riffe, D., & Lacy, S. (2019). *Analyzing media messages: Using quantitative content analysis in research*. Routledge.

Weihe, G. (2008). Public-Private Partnerships and Public-Private Value Trade-Offs. *Public Money & Management, June*, 153–158.

Weill, P., & Ross, J. (2005). A matrixed approach to designing IT governance. *MIT Sloan management review*, 46(2), 26.

Wende, K. (2007). A model for data governance - Organising accountabilities for data quality management. *ACIS 2007 Proceedings - 18th Australasian Conference on Information Systems*, 416–425.

Wendel, J., & Edberg, D. (2015). Outlook for state-designated health information exchange: Learning from the information systems and economics literatures. *Health Systems*, 4(1), 82–90.

Wendelborn, C., Anger, M., & Schickhardt, C. (2023). What is data stewardship? Towards a comprehensive understanding. *Journal of biomedical informatics*, 140, 104337.

Wet moderniserend elektronisch bestuurlijk verkeer. (2023, 10 mei). Geraadpleegd op 10 februari 2025, van <https://wetten.overheid.nl/BWBR0048252/2025-01-01>

Whiting, L. S. (2008). Semi-structured interviews: guidance for novice researchers. *Nursing Standard*, 22(23), 35–40.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of planning education and research*, 39(1), 93-112.

Yin, R. K. (2017). Case Study Research and Applications Design and Methods Sixth Edition. In *Thousand Oaks: Sage Publications, Inc.*

Zeng, Y., Zhang, Q., Zhao, Q., & Huang, H. (2023). Doing more among institutional boundaries: Platform-enabled government in China. *Review of Policy Research*, 40(3), 458–478.

Zhou, Z. H., Chawla, N. V., Jin, Y., & Williams, G. J. (2014). Big data opportunities and challenges: Discussions from data analytics perspectives [discussion forum]. *IEEE Computational intelligence magazine*, 9(4), 62-74.

Appendices

APPENDIX A: TABLE SLR

Table 23. Articles included in the literature review

Author	Title	Year
Lee T.Y., Yee P.K., Cheung D.W.	E-government data interoperability framework in Hong Kong	2009
Bharosa N., Janssen M., Hulstijn J., Van Wijk R., De Winne N., Tan Y.-H.	Towards a lean-government using new IT-architectures for compliance monitoring	2011
De Winne N., Janssen M., Bharosa N., Van Wijk R., Hulstijn J.	Transforming public-private networks: An XBRL-based infrastructure for transforming business-to-government information exchange	2011
He H., Zheng Y., Yang Y.	Research and design of heterogeneous data exchange system in e-government based on XML	2012
Klievink B., Janssen M., Tan Y.-H.	A stakeholder analysis of business-to-government information sharing: The governance of a public-private platform	2012
Klievink B., Janssen M., Tan Y.-H.	Blurring public-private boundaries: Governance of information sharing in global trade networks	2012
Bharosa N., Janssen M., van Wijk R., de Winne N., van der Voort H., Hulstijn J., Tan Y.-H.	Tapping into existing information flows: The transformation to compliance by design in business-to-government information exchange	2013
Bharosa N., Klievink B., Tan Y.-H., Janssen M.	Developing multi-sided platforms for public-private information sharing: Design observations from two case studies	2013
Janssen M., Tan Y.-H.	Dynamic capabilities for information sharing: XBRL enabling business-to-government information exchange	2014
Klievink B., Zuiderwijk A., Janssen M.	Interconnecting governments, businesses, and citizens - A comparison of two digital infrastructures	2014
Vest J.R., Campion T.R., Kern L.M., Kaushal R.	Public and private sector roles in health information technology policy: Insights from the implementation and operation of exchange efforts in the United States	2014
van Engelenburg S., Janssen M., Klievink B.	Design of a business-to-government information sharing architecture using business rules	2015
Wendel J., Edberg D.	Outlook for state-designated health information exchange: Learning from the information systems and economics literatures	2015
Luciano J.S., Sayogo D., Ran W., DePaula N., Jarman H., Tayi G., Zhang J., Hrdinova J., Pardo T., Andersen D.L., Andersen D.F., Luna-Reyes L.F.	Building a Certification and Inspection Data Infrastructure to Promote Transparent Markets	2017

Bharosa N., Hietbrink F., Mosterd L., Van Oosterhout R.	Steering the adoption of Standard Business Reporting for cross domain information exchange	2018
Paide K., Pappel I., Draheim D., Vainsalu H.	On the systematic exploitation of the Estonian data exchange layer X-road for strengthening public-private partnerships	2018
Van de Kaa G., Janssen M., Rezaei J.	Standards battle for business-to-government data exchange: Identifying success factors for standard dominance using the Best Worst Method	2018
Van Engelenburg S., Janssen M., Klievink B., Tan Y.-H., Rukanova B.	Comparing the openness of archetypical business-To-government information sharing architectures: Balancing advantages of openness with the control of risks	2018
Chen T., Yu Y., Duan Z	Data access sharing approach for trade documentations based on blockchain technology	2019
Hofman W.	Toward large-scale logistics interoperability based on an analysis of available open standards	2019
Kurnia R.A., Praditya D., Janssen M.	A Comparative Study of Business-to-Government Information Sharing Arrangements for Tax Reporting	2019
Mermoud A., Keupp M.M., Percia David D.	Governance models preferences for security information sharing: An institutional economics perspective for critical infrastructure protection	2019
Robinson N.J., Dahlhaus P.G., Wong M., MacLeod A., Jones D., Nicholson C.	Testing the public-private soil data and information sharing model for sustainable soil management outcomes	2019
Susha I., Rukanova B., Ramon Gil-Garcia J., Tan Y.-H., Hernandez M.G.	Identifying mechanisms for achieving voluntary data sharing in cross-sector partnerships for public good	2019
Rukanova B., Ubacht J., Van Engelenburg S., Tan Y.-H., Geurts M., Sies M., Molenhuis M., Slegt M.	Realising value from voluntary business-government information sharing through blockchain-enabled infrastructures: The case of importing tires to the Netherlands using TradeLens	2021
Aben T.A.E., van der Valk W., Roehrich J.K., Selviaridis K.	Managing information asymmetry in public-private relationships undergoing a digital transformation: the role of contractual and relational governance	2021
Zeng Y., Zhang Q., Zhao Q., Huang H.	Doing more among institutional boundaries: Platform-enabled government in China	2023

Table 24. Articles excluded due to inaccessibility

Author	Title	Year
Namdarian I.	The customs central applications platform for the government to government (g2g) and government to business (g2b) data exchange and retrieval	2011
Othman R.B., Mahamud K.R.B.K., Bakar M.S.B.A.	Public private collaboration model in spatial data infrastructure: Potential elements	2017
Phuaphanthong T., Bui T., Keretho S.	Establishing interagency collaboration in large-scale systems development: Lessons learned from an e-government project for trade and transport facilitation	2009
Vaidya K., Campbell J., Soar J., Gardner S.	Assimilation of inter-organizational information systems: Insight from change resistance theory in public electronic procurement	2011
Vos M., Cullen R., Cranefield J.	Model for privacy in public-private sector RFID systems	2014
Zavattero D.A.	D-Bay: A public-private data sharing straw-man" a project of the ITS America policy, evaluation, and advocacy forum	2006

APPENDIX B: CASE DESCRIPTIONS

CASE 1: BANK SUPERVISION

DNB supervises financial institutions in the Netherlands and ensures that they comply with their statutory reporting obligations. Banks, insurers, and pension funds must submit periodic financial reports, each under a different legal framework: banks report under CRD IV, insurers via data statements and pension funds under the Pension Funds Supervision Act. These reports are submitted to DNB and contain essential information on financial health, solvency, and risk management.

DNB collects and processes this data, checks whether institutions meet the reporting requirements, and in many cases forwards this information to European regulators. For banks, this is the European Banking Authority (EBA), and for insurers, the European Insurance and Occupational Pensions Authority (EIOPA). Sometimes this forwarding is done one-to-one, but often DNB also performs additional checks or edits before forwarding the data. This digital and standardised reporting chain contributes to effective financial supervision at the national and European levels and strengthens the stability and transparency of the European financial system.

CASE 2: BRP ISSUE REPORTING

Entities such as municipalities, educational institutions and healthcare providers that use national registry data in their operations are legally required to submit a report if they have reasonable grounds to doubt the accuracy of the data. This obligation is known as the 'feedback duty' (terugmeldplicht). For the national registry of persons (BRP), this is implemented through Terugmeldsysteem version 2.0 (TMV2.0), which is managed by the Dutch Identity Data Agency (RvIG). TMV2.0 uses a feedback functionality called Digimelding, which is a generic service within the Generic Digital Infrastructure (GDI) and is managed by Logius. Digimelding enables organisations to report suspected inaccuracies in data retrieved from national registries. Feedback can be submitted either via a portal, which requires authentication through eHerkenning, or directly via an organisation's own application, provided that their IT supplier has implemented the Digimelding web service. Digimelding plays a critical role in maintaining data quality within the BRP by enabling errors to be identified and corrected in a timely manner. This process is essential for ensuring the reliability of public services and government decision-making.

CASE 3: CHILD WELFARE SERVICES

In child welfare, standardised and digitised processes have been introduced. In 2023, the child welfare sector decided to enhance collaboration by introducing the concept of 'network cooperation at your fingertips'. This has four objectives: 1) reliable information for citizens; 2) insight and influence for citizens on their pathway; 3) better coordination between professionals; and 4) an up-to-date overview for professionals. Our example of this new type of PII is the authority process for the Marechaussee. It is crucial for these organisations to be able to determine whether a child is permitted to travel abroad. To achieve this, it is

necessary to consult not only the BRP, but also judicial decisions relating to custody and police records. Based on this data, it is crucial that they receive a definitive answer. This requires data exchange between many different parties, and additional services, such as permissions, are vital.

CASE 4: CARECODEX

BabyConnect is a national acceleration programme designed to enable the secure and user-friendly exchange of digital data within the birth care sector. Initiated by the CareCodex Foundation on behalf of the Ministry of Health, Welfare and Sport (VWS), it is implemented in collaboration with organisations such as RSO Netherlands, CPZ, Nictiz and Perined. BabyConnect aims to improve continuity of care by facilitating the digital exchange of information between various healthcare providers, including midwives, gynaecologists, maternity care professionals, youth health services, and paediatricians. At the core of the programme is the Birth Care Interoperability trust framework, which sets out the necessary agreements, architectural framework, and technical specifications to achieve interoperability in birth care. This approach reduces the risk of medical errors, lowers administrative burdens, and improves the quality of care overall. Furthermore, clients gain greater control over their personal health data, fostering transparency and active participation in the care process. Implementation takes place at the regional level through partnerships with obstetric care networks (VSVs) and other healthcare organisations.

CASE 5: ECODEx

In the Netherlands, eCodex is used, among other things, for cross-border judicial cooperation, such as the exchange of investigation information between Dutch and foreign authorities. The system contributes to more efficient and secure justice and security cooperation within the EU. eCodex is a European digital infrastructure that enables secure, cross-border communication between judicial authorities. The system supports both civil and criminal proceedings and enables competent authorities to exchange documents, legal forms, and evidence electronically. In the Netherlands, eCodex is managed by the Ministry of Justice and Security and is intended for use in proceedings with a European legal basis. The system was developed by 21 EU member states between 2010 and 2016 and is currently managed by a consortium of member states and other organisations. On December 2nd, 2020, the European Commission presented a legislative proposal to make eCodex permanent through an appropriate governance and management structure compatible with EU-LISA.

CASE 6: EDU-V

The Edu-V case concerns the standardised exchange of data for the procurement and access to digital learning materials in Dutch primary, secondary, special, and vocational education. It is supported by the Edu-V trust framework, which establishes a governance arrangement for collaboration between public and private stakeholders. The framework facilitates secure, reliable, and user-friendly digital data exchange within the education sector. It includes agreements on privacy, technical standards, reference components, and

data services/ Schools and suppliers develop these agreements jointly in co-creation. Within this PII, roles and responsibilities are clearly defined to ensure coordinated participation and accountability.

CASE 7: EINVOICING

Electronic invoicing to central government organisations in the Netherlands is handled via Digipoort, a digital hub that acts as a secure and standardised gateway for electronic invoices. Suppliers are required to submit invoices in UBL format according to the European standard EN16931. There are several main ways to send eInvoices to government bodies. Suppliers can submit eInvoices via direct connections, intermediating service providers or government supplier portals. At an international level, CEN working groups define European standards, while NEN working groups coordinate national governance. Additionally, there are several government-specific governance bodies for government-oriented invoice exchange. This case is presented in more detail in Chapter 4.

CASE 8: IMPORT DECLARATION

Businesses importing goods into the Netherlands are required to submit Customs declarations to the Dutch Customs Authority. These declarations are processed digitally via the 'Douane Aangifte Application' (DAA), a system designed to manage import data efficiently and securely. Each declaration includes key information such as the type, value, origin, and destination of the goods.

The digitisation of this process significantly reduces the administrative burden on companies by enabling streamlined online submission and faster clearance of goods. The system performs automatic checks and calculates applicable taxes and import duties in real time, ensuring accurate revenue collection and prompt shipment release.

CASE 9: KIK-V

In the case of KIK-V, a PII is installed to enable healthcare providers, regulators, and policymakers to have insight into the business administration of nursing homes. This case uses the KIK-V trust framework. The trust framework contains agreements on data standards, data definitions, and the safe exchange of information. It uses existing standards such as iStandards and ZIBs (healthcare information building blocks). The focus is on reusing data already available at healthcare organisations. The aim is to reduce the administrative burden and improve the quality of care information. KIK-V enables healthcare providers, regulators, and policymakers to make better use of existing data. There is a temporary central system built in the case of KIK-v to enable information obtaining organisations to use the network without initial implementation. The objective is to create a distributed network, in which each data sender and receiver has their own system, which connects peer-to-peer with other senders or receivers, based on the agreements made in the trust framework.

CASE 10: LAND REGISTRY

In the Netherlands, notaries play a central role in the legal transfer of property ownership. They are responsible for drafting and registering the purchase agreement and deed of transfer in the Land Registry (Kadaster), which serves as the official record of property rights. This registration provides legal certainty and is essential for safeguarding property ownership. Through ongoing digitisation, systems like Kadaster Online enable swift and secure access to ownership data for both professionals and the general public. This digital infrastructure streamlines the registration and verification process for deeds and mortgages, reduces administrative burdens, and lowers the risk of errors. The public nature of the register enhances transparency and helps to prevent legal conflicts. The digitisation of property registration thus contributes to a more efficient, reliable, and transparent housing market.

CASE 11: MIGRATION

The Ministry of Justice and Security (Min J&V) acquires information from the partners in the chain, for example, from the immigration and naturalisation service. Min J&V integrates this information into the form of information products, which are then used to serve the (governmental) stakeholders. For instance, the minister, the House of Representatives, European regulators, or the press. This is predominantly achieved through the utilisation of aggregated reports. They refrain from commenting on individuals and solely provide insights into trends and developments.

CASE 12: NATIONAL HOUSING PROGRESS

In the Netherlands, housing construction is monitored at the municipal and provincial levels. These organisations have to provide data to the LMVB. The LMVB is a data facility that brings together information on planned capacity, granted permits, and completed projects to monitor the progress of housing construction at a national level. During the development of this monitor, uniform definitions were established for all levels of government and set out in the LMVB Data Protocol. Municipalities provide data to provincial plan monitors via tools such as DIWI, which they then contribute to the LMVB. DIWI is a data tool developed by the VNG in cooperation with municipalities to improve the internal provision of information on housing projects. It enables municipalities to store and manage building plan data in a structured way, providing greater insight into progress and simplifying communication. DIWI improves data quality, reduces errors in data exchange, and promotes a data-driven approach within municipalities. The VNG provides project management and support for the implementation of DIWI.

CASE 13: NOTIFICATION PROPERTY VALUE

In the case of property value notifications, the National Land Registry Organisation sends updates of digital data on the assessed value of properties, such as homes and commercial buildings, to municipalities and other government bodies once a year. Using a digital notification service means that property value data which has changed over the year can be sent quickly, safely, and uniformly to the relevant parties, such as

the tax authorities. Digitising the property value notification process offers benefits such as reduced administrative burden for municipalities, faster processing of data, and higher-quality information. This reliable data sharing ensures that tax levies based on property values are accurate. This notification service is offered by a central government organisation and can be used for reusing property data in other types of data exchange.

CASE 14: OOTS

Within Europe, a coordinated framework exists to facilitate the exchange and verification of diploma information between national authorities. When a student applies for further education abroad—for example, a master’s programme—educational institutions and competent authorities can communicate through a shared digital infrastructure. One organisation can notify another that it has received a diploma copy from a student, typically in the form of a scanned PDF. Through a central messaging system, this notification is shared with the relevant national authority in the country of origin. Data can, with the consent of the data subject, be shared by government organisations across the EU. This is done by the system (once only technical system), OOTS. OOTS aims to reduce the administrative burden on citizens and businesses and improve the efficiency of public services.

CASE 15: PENSION FUNDS

The administration of pension funds is the responsibility of pension administration organisations. This is done in accordance with the prevailing pension regulations. To administer the pension fund in an appropriate manner, it is necessary to obtain data from external companies. For instance, an individual who is entitled to a pension or is retired may fall ill and become disabled or unemployed. To proceed with this matter, it is necessary that pension funds obtain relevant data from UWV. These are received monthly.

CASE 16: SBR-EDUCATION

In the Netherlands, educational institutions are legally required to submit their annual financial statements to DUO (the Netherlands Organisation for Student Affairs), thereby providing transparency about their financial position and enabling accountability to the government. This reporting obligation applies to primary and secondary schools alike.

Since 2015, these statements have had to be submitted digitally by the 1st of July via the XBRL Education Portal. XBRL (eXtensible Business Reporting Language) is an internationally recognised standard for the digital exchange of financial data. Using XBRL streamlines the reporting process, enabling institutions to submit data efficiently and allowing DUO to automate validation and processing.

DUO operates under the broader SBR (Standard Business Reporting) governance framework, which facilitates the standardisation and reuse of reporting components across sectors. This system uses a central infrastructural component called Digipoort. This system is shared by multiple public organisations and is provided by the public intermediary Logius.

CASE 17: SINGLE WINDOW

The Single Window for Customs is a digital platform designed to streamline communication between businesses and the relevant government authorities involved in importing, exporting and transiting goods. In the Netherlands, the system is managed by Customs and is part of the wider EU Single Window Environment for Customs, which was set up under EU Regulation 2022/2399 in December 2022. The platform enables companies to submit all the necessary data and documentation for cross-border trade electronically and centrally. In the Dutch context, the Single Window is used in various sectors, including air and maritime transport. By improving coordination between Customs and other relevant authorities, such as the Royal Military Police (KMar) and the Netherlands Food and Consumer Product Safety Authority (NVWA), the system helps create a more efficient, secure, and transparent trading environment. Businesses benefit from a simplified process where information only needs to be provided once, reducing the risk of errors and shortening processing times.

CASE 18: VACCINATION DATA

The National Switch Point (LSP) facilitates digital data exchange between Youth Health Care (JGZ) organisations and the National Institute for Public Health and the Environment (RIVM). This infrastructure ensures the secure and efficient transfer of vaccination data. After vaccinations have been given, the data is provided from JGZ and sent through the LSP to the RIVM. The Association of Healthcare Communication Providers (VZVZ) is the intermediary organisation and provides the LSP. They collaborate with organisations such as NICTIZ and HL7 to develop data formats and standards for healthcare.

CASE 19: VUM

The aim of the Improving Matching Data Exchange (VUM) programme is to enhance data exchange between public and private organisations within the work and income domain. It aims to improve and accelerate the matching of job seekers to suitable vacancies by promoting standardised, secure data sharing. VUM is a collaborative initiative involving the Association of Netherlands Municipalities (VNG), the Employee Insurance Agency (UWV) and the Ministry of Social Affairs and Employment (SZW). The programme uses standardised datasets and modern technologies, such as API integrations, to enable interoperability. By increasing transparency across regional and organisational boundaries, VUM enables more effective and efficient labour market matching.

CASE 20: WAGE RETURN STATISTICS

Statistics Netherlands (CBS) is a key player in the wage declaration chain, a partnership in which data from wage declarations are collected centrally and reused for various public purposes. The wage declaration chain consists of the Tax Administration, the unemployment office (UWV) and the statistics agency (CBS), and makes it possible to obtain up-to-date and reliable information on work, income, and labour relations.

Employers submit wage declarations to the Tax and Customs Administration, which then makes this data available to UWV and CBS.

CBS uses these data for official statistics and policy analyses, such as on employment, income distribution, and social security. Using existing declarations avoids double registration and reduces administrative burdens. The wage declaration chain is based on uniform definitions and data quality requirements, so that data from different sources are consistent and comparable.

CASE 21: WMO

The iWmo standard enables secure and consistent digital communication between municipalities and care providers operating under the Social Support Act (Wmo). Developed by Zorginstituut Nederland in collaboration with sector stakeholders, it ensures consistent data exchange throughout the care process, from initial assessment to final declaration. It includes structured message types such as the WMO301 for allocating support and the WMO305 for commencing care services. This standardised approach enhances collaboration between municipalities and care providers, minimises administrative errors and reduces the bureaucratic workload. The technical infrastructure supporting this messaging system is maintained by organisations such as VECOZO, who ensure secure and efficient data processing. By digitising these processes, the iWmo standard contributes to making the Wmo system more transparent, efficient, and client-focused, enabling support to be delivered faster and more appropriately.

CASE 22: WOCO

Housing associations in the Netherlands are required to submit two digital reports annually: the dVi (Digital Accountability Information) and the dPi (Digital Forecast Information). The dVi reflects past performance, while the dPi outlines the multi-year budget and financial projections. Both reports must be submitted by 30 June, either via manual input or through Standard Business Reporting (SBR) using system-to-system (S2S) connections from internal software.

Digitising these processes reduces administrative burden, enhances data quality, and enables more effective supervision and transparency within the sector. This reporting framework forms part of the Improving the Provision of Information in the Housing Corporation Sector covenant and contributes to a more standardised and efficient exchange of information between stakeholders.

Summary

Governments rely heavily on information to perform their public duties. Much of the relevant data is held by other parties, including citizens, private companies, knowledge institutes, civil society organisations, and (international) data platforms. Over the past decade, we have seen a rise in public information infrastructures (PII) to support data exchange between these (external) parties and public organisations. PIIs are ill understood, and public agencies looking for guidance on designing PII will find that insights in the literature are scarce and scattered. The main objective of this dissertation is to provide academically grounded and empirically driven knowledge on PII configurations and stewardship. Gaining insight into the configuration of PII and how these can be steered by PII stewards early in the process can help public organisations make more informed, coherent decisions and ensure the infrastructure evolves in alignment with public values and long-term goals. The main question of this research is two-fold: **What is PII stewardship, and how does PII stewardship relate to configurations for data exchange with (semi-) public organisations?** Through a series of research questions, we have examined PII stewardship in terms of the quality measures of a PII, design variables in a PII, the prevailing configurations of PII, and the different categories of PII Stewardship. These questions have been divided into four phases, as shown in Table 25.

Table 25. Research phases and sub-research questions.

Phase	Research question
1. Definitions	What is a public information infrastructure?
	How can public information infrastructure stewardship be defined?
2. Design choices	What are the relevant design choices in public information infrastructures?
3. Configurations	What are the prevailing configurations of public information infrastructures through which organisations exchange data with (semi-)public organisations?
	What are the strengths and weaknesses of the prevailing configurations of public information infrastructures in terms of their quality?
4. Stewardship	What classification of public information infrastructure stewardship can be distilled from the findings of this study?

In phase 1: *Definitions*, this research started with a literature review of the concepts of PII and of stewardship. Here, we established that PIIs are a specific form of information infrastructures. We define PII as socio-technical systems composed of interrelated technical and social components, which enable the exchange of information with public organisations. The PII configuration can be defined in terms of data, technical components, and governance configurations. PII evolve over time, rather than emerging spontaneously, and are shaped by the decisions and interactions of the involved actors.

To shape and steer the PII, we defined a new concept: Public Information Infrastructure Stewardship. PII-stewardship refers to the authorities for maintaining, adapting and evolving the information infrastructure throughout its various life cycles, for ensuring data, technical and governance quality. Based on the literature, we established that there are multiple categories of stewardship possible, including at least

distributed stewardship, next to single-actor stewardship. Due to the fact that multiple actors can play a role in a PII, collaboration can be sought between the different stakeholders for stewardship roles, while other PII's may steer toward single-actor stewardship.

These PII stewards must steer for high-quality PII. There is, however, limited research defining PII quality measures. Based on the literature, we established a framework for PII quality, providing the different measures of data quality, technical quality, and governance quality. This framework is validated through interviews with respondents from three case studies (Standard Business Reporting, Single Window, eInvoicing).

In phase 2: *Design choices*, we studied the relevant design choices PII stewards can make to configure a PII. We started with a literature review using the PRISMA method. This was followed by a multiple-case study approach, encompassing three case studies (Standard Business Reporting, Single Window, eInvoicing), based on a combination of interviews and document analysis. We established a total of twenty design variables divided into three categories: data, technical, and governance design choices. Our analysis identified ten relevant data design variables, eleven relevant technical design variables (covering systems, standards, and services), and nine governance design variables.

In phase 3: *Configurations*, the objective was to find prevailing configurations of PII design choices and research the strengths and weaknesses of these configurations. In this phase, the configurations of twenty-two cases of PII in the Dutch context were explored through a multiple-case study approach, based on data collected through interviews, surveys, and document review. Preliminary analysis included an examination of sub-configurations for the design choices in the data, governance, and technical components. A hierarchical clustering shows that a variety of data, governance, and technical sub-configurations are present within the overall PII configurations. Based on these sub-configurations, we then identified six prevailing configurations. These configurations include (1) Central Data Reuse PII, (2) EU-Regulation PII, (3) Technical-Driven PII, (4) Central Public Collaboration PII, (5) Mixed Collaboration PII, and (6) Distributed PII with a Trust Framework.

The first configuration, the central data reuse PII, is characterised by a centrally organised system that involves minimal cooperation between involved parties. The second configuration, the EU-regulated PII, is also characterised by a centrally organised system with limited cooperation between the relevant actors. Yet here, the initiation of data exchange is not set at the national level, but rather at the European level, usually through EU regulations or directives. The third configuration is the technical-driven PII, which originates from European initiatives that are primarily technical in nature. These initiatives focus on establishing technical exchange mechanisms for cross-border data exchange and harmonising exchange specifications, such as transport protocols and authentication. Configuration four, the public collaboration PII, is characterised by centralised cooperation, where government-wide standards play a central role. Configuration five, the mixed collaboration PII, is characterised by a hybrid information infrastructure that integrates both centralised systems and decentralised or distributed components. Finally, configuration six, the distributed PII, is distinguished by the distributed nature of the information infrastructure and the focus on trust frameworks as a coordination mechanism.

For the purpose of the study, a selection of respondents per case were requested to forward a survey to three relevant actors involved in the same PII. The survey asked respondents about their perceptions of the PII's strengths and weaknesses in data quality, technical quality, and governance quality. We observe that these configurations have different strengths and weaknesses regarding data, technical and governance quality. It is essential to note that no particular infrastructure configuration is universally superior to others. Each of these approaches has its own strengths and weaknesses. PPI stewards must be aware of these characteristics in order to deal with them accordingly.

Finally, in phase 4: *Stewardship*, the last phase of this research, we synthesised the acquired knowledge to identify and understand PII stewardship in relation to the configurations. This enabled us to enrich the conceptual analysis of PII stewardship and empirically derive to a classification of PII stewardship. We find that the classification of PII stewardship consists of three categories: (1) single-actor stewardship, (2) distributed stewardship, and (3) shared stewardship. First, in single actor stewardship, as the term suggests, there is a single authority for steering the entire PII. We found, in the cases, that single-actor stewardship is primarily based on laws and regulations and can encourage growth and utilisation in the early stages of a PII. The second category of stewardship is distributed PII stewardship. Here, what may begin as a regulatory requirement can evolve into a more cooperative approach, with authorities actively seeking alignment. By collaborating, using standards, and decentralising the PII stewardship, it becomes distributed across multiple authorities. This makes alignment between the different parts of the PII more complex and highlights the need for continuous coordination among the participating parties. The third and last category of stewardship identified is shared PII stewardship. Shared PII stewardship represents the highest degree of centralised cooperation. Here, multiple authorities collaboratively form and shape the PII in a single governance structure. In this form, stewardship is shared among a collective of parties, often supported by a coordinating organisation or programme structure.

Based on the research findings, the answer to the first part of the main question – What is PII stewardship? – is formulated as follows: *PII stewardship refers to the authorities for ensuring data, technical and governance quality in designing, maintaining, and adapting a public information infrastructure throughout its life cycles.* The answer to the second part of the main question – How does PII stewardship relate to configurations for data exchange with (semi-) public organisations? – entails the relation between the three types of stewardship as discussed above and the six prevailing configurations. First of all, single-actor PII stewardship often relates to configurations such as central data reuse PII and EU-regulated PII. Secondly, distributed PII stewardship tends to relate to configurations based on either a mixed collaboration or public collaboration PII. Finally, shared PII stewardship often relates to distributed PII configuration or the technical-driven configuration.

By answering the main question, this study contributes to both academic and practical discussions on the design and stewardship of PII. This dissertation contributes to the academic understanding of public infrastructure governance by introducing a typology of six infrastructural configurations and providing the first overview of PII sub-configurations and prevailing PII configurations. It thereby establishes a foundational understanding of the patterns in which these configurations appear. By identifying the design

choices that form PII configurations, this dissertation paves the way for future studies to develop testable hypotheses and provides a basis for further theorisation of PII. Moreover, highlighting the alignment problem faced by PII stewards in PII configurations adds a new layer to the existing literature by framing alignment not as a design-time challenge but as an ongoing institutional and technical effort that persists across the lifecycle of digital infrastructures.

Additionally, this dissertation introduces the novel concept of PII stewardship. Drawing on existing literature and the complexities of practice, we have identified the need for a clear and comprehensive definition of PII stewardship. As public agencies become more reliant on (external) data and socio-technical components beyond their direct or sole control, their ability to exercise adequate PII stewardship is essential for safeguarding public values. For practitioners such as policymakers, public managers, and infrastructure designers, this study contributes by offering insights into the path dependencies that follow from PII choices, providing design configurations for comparison, and offering insight into the various types of PII stewardship and their strengths and weaknesses. Further research avenues include examining the private perspective on (public) information infrastructures and comparing PII configurations with private information infrastructures. Furthermore, it is possible to replicate this research with additional cases of PII and PII stewardship from other countries to either confirm or supplement the design variable model.

Samenvatting

Overheden zijn sterk afhankelijk van informatie om hun publieke taken uit te voeren. Veel van de relevante gegevens zijn in handen van derde partijen, waaronder burgers, particuliere bedrijven, onderzoeksinstellingen, maatschappelijke organisaties en internationale platforms. De afgelopen jaren hebben we een toename gezien van publieke informatie-infrastructuren (PII's) ter ondersteuning van de gegevensuitwisseling tussen deze (externe) partijen en publieke organisaties. Zowel in de wetenschap als in de praktijk is er een gebrek aan inzicht in deze PII, waardoor het moeilijk is om te anticiperen op hoe publieke professionals, of in het bijzonder PII-Stewards, kunnen sturen op configuraties van hoge kwaliteit.

Bestaande literatuur biedt beperkt inzicht in configuraties van PII's, het ontwerpen van kwalitatieve PII en PII Stewardship. In een vroeg stadium inzicht verwerven in de configuratie van PII en hoe deze gestuurd kunnen worden door PII-stewards, kan overheidsorganisaties helpen beter geïnformeerde, coherente beslissingen te nemen en ervoor te zorgen dat de infrastructuur evolueert in overeenstemming met publieke waarden en lange termijn doelstellingen

De hoofdvraag van dit onderzoek is tweeledig: wat is PII-stewardship en hoe verhoudt PII-stewardship zich tot configuraties voor gegevensuitwisseling met (semi-)overheidsorganisaties? Door middel van een reeks onderzoeksvragen hebben we PII-stewardship onderzocht in termen van de kwaliteitsmaatstaven van een PII, ontwerpvariabelen in een PII, de gangbare configuraties van een PII en de verschillende soorten PII-Stewardship. Deze vragen zijn onderverdeeld in vier fasen, zoals weergegeven in tabel 26.

Tabel 26. Onderzoeksfases en onderzoeksvragen.

Fase	Research question
1. Definities	Wat is een publieke informatie infrastructuur?
	Hoe kunnen we publieke informatie infrastructuur stewardship definiëren?
2. Design variabelen	Wat zijn relevante ontwerpvariabelen in publieke informatie infrastructuren?
3. Configuraties	Wat zijn de configuraties van publieke informatie infrastructuren voor data uitwisseling met (semi-)publieke organisaties?
	Wat zijn de sterktes en zwaktes van de configuraties in termen van de PII-kwaliteit?
4. Stewardship	Welke classificatie van publieke informatie-infrastructuurstrating kan uit de bevindingen van deze studie worden afgeleid?

In fase één, genaamd *Definities*, begint dit onderzoek met een literatuuronderzoek naar de concepten van publieke informatie-infrastructuren en het beheer van publieke informatie-infrastructuren. We hebben vastgesteld dat PII een specifieke vorm van informatie-infrastructuur is. Publieke informatie-infrastructuren zijn socio-technische systemen die bestaan uit onderling verbonden technische en sociale componenten, die de uitwisseling van informatie voor publieke organisaties mogelijk maken. De PII-configuratie kan worden gedefinieerd in termen van gegevens-, techniek- en governance configuraties. De infrastructuren evolueren

in de loop van de tijd en ontstaan niet spontaan, maar worden gevormd door de beslissingen en interacties van de betrokken actoren.

Om de PII vorm te geven en te sturen, definiëren we een nieuw concept: publieke informatie-infrastructuur stewardship. PII-stewardship verwijst naar de autoriteiten met bevoegdheden voor het ontwerpen, onderhouden en aanpassen van de informatie-infrastructuur gedurende de verschillende levenscycli ervan, om de kwaliteit van de gegevens, de techniek en het beheer te waarborgen.

Op basis van de literatuur hebben we vastgesteld dat er meerdere soorten stewardship mogelijk zijn, waaronder gedeelde stewardship en stewardship door één enkele actor. Omdat meerdere actoren een rol kunnen spelen in een PII, kan samenwerking worden gezocht tussen de verschillende belanghebbenden voor stewardship-rollen, terwijl andere PII's kunnen neigen naar beheer door één enkele actor.

Deze PII-beheerders moeten streven naar PII van hoge kwaliteit. Er is echter weinig onderzoek gedaan naar het definiëren van kwaliteitsmaatstaven voor PII. Op basis van de literatuur hebben we een kader voor PII-kwaliteit opgesteld, met verschillende maatstaven voor data-, technische-, en governance-kwaliteit. Dit kader is gevalideerd door middel van interviews met respondenten uit drie casestudy's (Standard Business Reporting, Single Window, eInvoicing).

In de tweede fase, genaamd *Ontwerp variabelen*, hebben we gekeken naar de relevante ontwerpkeuzes die PII-Stewards kunnen maken om een PII te configureren. We zijn begonnen met een literatuuronderzoek volgens de PRISMA-methode. Daarna hebben we een meervoudige casestudybenadering gevolgd, bestaande uit drie casestudy's (Standard Business Reporting, Single Window, eInvoicing), gebaseerd op een combinatie van interviews en documentanalyse. We hebben in totaal twintig ontwerpvariabelen vastgesteld, onderverdeeld in drie categorieën: ontwerpkeuzes op het gebied van data, techniek en governance. Onze analyse identificeerde tien relevante gegevens ontwerpvariabelen, elf relevante technische ontwerpvariabelen, waaronder systemen, standaarden en services, en negen governance-ontwerpvariabelen.

In de derde fase, genaamd *Configuraties*, was het doel om gangbare configuraties van PII-ontwerpkeuzes te vinden en de sterke en zwakke punten van deze configuraties te onderzoeken. In de derde fase van dit onderzoek worden de (sub-)configuraties van tweeëntwintig casussen van PII in de Nederlandse context onderzocht door middel van een meervoudige casestudy waarbij interviews, enquêtes en documenten worden geanalyseerd. De eerste analyse omvat een onderzoek naar sub-configuraties voor de ontwerpkeuzes met betrekking tot de gegevens, de governance en de technische componenten. Een hiërarchische clustering toont aan dat er binnen de algemene PII-configuraties een verscheidenheid aan gegevens-, beheer- en technische sub-configuraties aanwezig is. Op basis van de clustering hebben we vervolgens zes overheersende configuraties geïdentificeerd. Deze configuraties zijn: PII met centraal hergebruik van gegevens, EU-regelgeving gebaseerde PII, technisch aangestuurd PII, centrale openbare samenwerking PII, gemengde samenwerking PII en gedistribueerd PII met een afsprakenstelsel.

De eerste configuratie, centraal gegevenshergebruik PII, wordt gekenmerkt door een centraal georganiseerd systeem met minimale samenwerking tussen de betrokken partijen. De tweede configuratie, de door de EU gereguleerde PII, wordt ook gekenmerkt door een centraal georganiseerd systeem met beperkte samenwerking tussen de relevante actoren. Toch wordt het initiatief voor gegevensuitwisseling hier

niet op nationaal niveau genomen, maar op Europees niveau, meestal door middel van EU-verordeningen of -richtlijnen. De derde configuratie, de technische PII, komt voort uit Europese initiatieven die vooral technisch van aard zijn. Deze initiatieven richten zich op het opzetten van technische uitwisselingsmechanismen voor grensoverschrijdende gegevensuitwisseling en het harmoniseren van uitwisselingsstructuren. Configuratie vier, de centrale overheidssamenwerking PII, wordt gekenmerkt door gecentraliseerde samenwerking op het gebied van digitale systemen, waarbij de overheid brede standaarden een centrale rol speelt. Configuratie vijf, de gemengde samenwerking PII, wordt gekenmerkt door een hybride informatie-infrastructuur die zowel gecentraliseerde systemen als gedecentraliseerde of gedistribueerde componenten integreert. Configuratie zes, de gedistribueerde PII, onderscheidt zich door de gedistribueerde aard van de informatie-infrastructuur en de daarmee samenhangende sterke focus op vertrouwenskaders als coördinatiemechanisme.

Voor het onderzoek is een selectie van respondenten gevraagd om een enquête door te sturen naar drie relevante actoren die betrokken zijn bij de PII. In de enquête is gevraagd naar de sterke en zwakke punten op het gebied van datakwaliteit, technische kwaliteit en governance kwaliteit in de PII. De zes configuraties hebben verschillende sterke en zwakke punten met betrekking tot de kwaliteit van gegevens, techniek en governance. Het is belangrijk op te merken dat geen enkel type infrastructuur universeel superieur is. Het is de verantwoordelijkheid van PII-stewards om zich bewust te zijn van de kenmerken van de configuraties en er dienovereenkomstig mee om te gaan.

Tot slot voeren we in fase 4, genaamd *Stewardship*, een synthese uit om de verworven kennis over de stewardship-rollen die bij de configuraties horen te identificeren en te begrijpen. Dit stelt ons in staat om de conceptuele analyse van PII-stewardship te verrijken en de soorten PII-stewardship in de praktijk in kaart te brengen. Door een synthese van de eerdere onderzoeksvragen te maken, creëren we een overzicht van de verschillende soorten PII-stewardship, waaronder single actor stewardship, gedistribueerd stewardship en shared stewardship. Ten eerste is er bij single actor stewardship, zoals de term al aangeeft, één enkele autoriteit die de PII beheert. We zien dat single actor stewardship voornamelijk gebaseerd is op wet- en regelgeving en voornamelijk in het begin van een PII kan stimuleren tot groei en gebruik. Het tweede type stewardship is gedistribueerd stewardship. Wat kan beginnen als een wettelijke vereiste kan zich ontwikkelen tot een meer coöperatieve aanpak, waarbij partijen actief op zoek gaan naar afstemming. Door samen te werken, standaarden te gebruiken en te decentraliseren, wordt het beheer van de PII gedistribueerd over meerdere actoren op meerdere plekken binnen de PII. Dit maakt handhaving complexer en benadrukt de noodzaak van voortdurende coördinatie en afstemming van belangen tussen de deelnemende partijen. Het derde en laatste type stewardship is gedeeld stewardship van PII. Gedeeld PII-stewardship vertegenwoordigt de hoogste mate van gecentraliseerde samenwerking en gedeelde verantwoordelijkheid die in de praktijk binnen PII wordt aangetroffen. In deze vorm wordt het stewardship gedeeld door een collectief van autoriteiten, dat vaak wordt ondersteund door een coördinerende organisatie of programmastructuur.

Op basis van de onderzoeksresultaten wordt het antwoord op het eerste deel van de hoofdvraag – *Wat is PII-Stewardship?* – als volgt geformuleerd: PII-Stewardship verwijst naar de autoriteiten voor het ontwerpen, onderhouden en aanpassen van de informatie-infrastructuur gedurende de gehele levenscyclus

ervan, met als doel de kwaliteit van de gegevens, de techniek en de governance te waarborgen. Het antwoord op het tweede deel van de hoofdvraag – *Hoe verboudt PII-stewardship zich tot configuraties voor gegevensuitwisseling met (semi-)publieke organisaties?* – omvat de drie hierboven besproken soorten PII-stewardship: (1) stewardship door één actor, (2) gedistribueerde stewardship en (3) gedeelde stewardship.

Door de hoofdvraag te beantwoorden, draagt dit onderzoek bij aan zowel academische als praktische discussies over het ontwerp en beheer van PII. Dit proefschrift draagt bij aan het academische begrip van het beheer van publieke informatie infrastructuren door een typologie van zes infrastructurele configuraties te introduceren en het eerste overzicht te geven van PII sub-configuraties en gangbare PII-configuraties. Het legt daarmee een fundamentele basis voor het begrip van de patronen waarin deze configuraties voorkomen. Door als eerste de ontwerpkeuzes te identificeren die PII-configuraties vormen, maakt dit proefschrift de weg vrij voor toekomstige studies om toetsbare hypothesen te ontwikkelen en biedt het een basis voor verdere theoretisering van PII. Bovendien voegt het benadrukken van het afstemmingsprobleem waarmee PII-Stewards in PII-configuraties worden geconfronteerd, een nieuwe laag toe aan de bestaande literatuur door afstemming niet te beschouwen als een uitdaging op het moment van ontwerp, maar als een voortdurende institutionele en technische inspanning die gedurende de hele levenscyclus van digitale infrastructuren blijft bestaan.

Daarnaast introduceert dit proefschrift het nieuwe concept van PII-Stewardship. Op basis van bestaande literatuur en complexiteiten in de praktijk hebben we vastgesteld dat er behoefte is aan een duidelijke en uitgebreide definitie van PII-Stewardship. Naarmate overheidsinstanties steeds meer afhankelijk worden van (externe) gegevens en sociaal-technische componenten waarover zij geen directe of volledige controle hebben, is hun vermogen om adequaat PII-stewardship uit te oefenen essentieel voor het waarborgen van publieke waarden. Voor praktijkmensen zoals beleidsmakers, openbare managers en infrastructuurontwerpers levert deze studie een bijdrage door inzicht te bieden in de pad afhankelijkheden die voortvloeien uit PII-keuzes, door verschillende ontwerpconfiguraties ter vergelijking aan te reiken en door inzicht te bieden in de verschillende soorten PII-stewardship en hun sterke en zwakke punten. Verdere onderzoeksmogelijkheden zijn onder meer het onderzoeken van het particuliere perspectief op (publieke) informatie-infrastructuren en het vergelijken van PII-configuraties met private informatie-infrastructuren. Bovendien is het mogelijk om dit onderzoek te repliceren met aanvullende gevallen van PII en PII-stewardship uit andere landen om het ontwerpvariabelenmodel te bevestigen of aan te vullen.

Curriculum Vitae

Wendy van Donge was born in Heemskerk on June 6, 1991. She has a broad academic background and extensive experience in the public sector, with a focus on data-driven government processes, standardisation, and program management. She began her academic journey at Leiden University (2011-2014), where she earned a Bachelor's degree in Public Administration, specialising in Economics, Politics, and Management. During her time at Leiden, she developed a strong affinity for the intersection of public administration and technology. She then pursued a Research master's in Public Administration and Organisational Science at Utrecht University (2014-2017). Her master's thesis focused on modelling the public sector using Discrete Simulation Modelling (DSM) to analyse throughput times of court cases within the Dutch justice system, demonstrating how DSM can be applied in seemingly non-technical environments and the value it can bring to such complex socio-technical systems.

Since 2018, Wendy has been working at Logius. Logius is the digital government service of the Netherlands Ministry of the Interior and Kingdom Relations (BZK). She started as Program Manager for Digipoort, responsible for multi-year budgeting, impact analyses, stakeholder coordination, and drafting business cases. Later, she transitioned to a consultant role for the Digipoort tendering process, where she wrote documentation and provided technical content. Upon completion of this program, she was appointed as a Standardisation Advisor in the e-procurement domain. In this role, she continues to contribute to the development and ongoing improvement of standards and participates in national and European working groups.

Since 2020, alongside her work at Logius, Wendy has been affiliated with Delft University of Technology, where she continues to deepen her expertise, has published several scientific articles, and has supervised graduate theses and guided literature research for six students during their courses. Her research focuses on data-driven government and information infrastructures, including publications on data stewardship and the role of artificial intelligence in public organisations. Her ongoing manuscript examines the connection between configurations of public information infrastructures and stewardship through a multiple-case analysis.


With the government increasingly combining practical experience with academic knowledge and evidence-based approaches, Wendy's capabilities can contribute to well-founded progress in public-sector infrastructure.

Publications:

- Van Donge, W., Bharosa, N., & Janssen, M. (2020). Future government data strategies: data-driven enterprise or data steward? Exploring definitions and challenges for the government as a data

enterprise. In *Proceedings of the 21st Annual International Conference on Digital Government Research* (pp. 196-204).

- Van Donge, W., Bharosa, N., & Janssen, M. (2022). Data-driven government: Cross-case comparison of data stewardship in data ecosystems. *Government Information Quarterly*, 39(2), 101642.
- Van Donge, W., Janssen, M. F. W. H. A., & Bharosa, N. (2024, August). Preparing Public Agencies for Harnessing AI: A Study on Variables Shaping Multi Actor Information Infrastructures. In *International Conference on Electronic Government* (pp. 254-269). Cham: Springer Nature Switzerland.
- Van der Peet, L., van Donge, W., Bharosa, N., & Janssen, M. (2025, August). Designing for Trust in Healthcare Data Sharing: Trust Anchors in the Trust Framework Lifecycle. In *International Conference on Electronic Government* (pp. 150-163). Cham: Springer Nature Switzerland.



Governments rely heavily on information to perform their public duties. Much of the relevant data is held by other parties, including citizens, private companies, knowledge institutes, civil society organisations, and data platforms. Over the past decade, we have seen a rise in public information infrastructures (PII) to support data exchange between these parties and public organisations.

This thesis provides academically grounded and empirically driven knowledge on PII configurations and stewardship. The development of public information infrastructures requires more than the design of temporary technical solutions or one-off institutional arrangements. They demand continuous alignment efforts and dedicated PII stewardship to ensure that infrastructures evolve with their context, maintain their quality, and remain fit for public purpose.

While there is no universal blueprint, the prevailing configurations and categories of PII stewardship offered in this manuscript provide a foundation for both scholars and practitioners to better understand and shape these information infrastructures in an increasingly digital society.